November 2020

Within-Field Yield Prediction for Sugarcane and Rice Focused on Precision Agriculture Applications

Felippe Hoffmann Silva Karp

WITHIN-FIELD YIELD PREDICTION FOR SUGARCANE AND RICE FOCUSED ON PRECISION AGRICULTURE APPLICATIONS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The School of Plant, Environmental, & Soil Sciences

by
Felippe Hoffmann Silva Karp
B.S., University of Sao Paulo, 2018
December 2020

To my parents Willian Feldberg Karp and Paula Emilia Hoffmann Silva Karp *(in memoriam)* who always gave their best when raising me and my sister. It is because of their effort and dedication that I am here today. Love you both.

Acknowledgments

First, I am grateful to God who gave me the strength and knowledge to be here today. I am also grateful for my wife Karen Florentino Correa, who was by my side giving me strength, support, and a lot of love while I was developing this project. I could not have someone better than her by my side. I also thank Darcy and Ivone, who always gave us a lot of support and allowed their beloved daughter to marry me and come with me to the US. A big thanks goes to my family, my dad, my sister Aline, my grandmothers Paulina (*in memoriam*) and Neuza, my aunts Paulette and Maide. People who would call me every week just to check on me and see if I needed anything. They are awesome.

I would also like to thank my American family, The Kerrs. Thank you, mom, Beth Kerr, dad, Scott Kerr, and my brothers: Jacob, Clayton, Fallon, and Cameron. Love you all so much. Nothing pays for all the love you all shared with me.

I would like to show my appreciation for my major advisor, Dr. Luciano Shiratsuchi, who first gave me the opportunity to join LSU and become one of his master's students. He also allowed me to have a lot of experiences throughout these two years. The knowledge exchanged, the nice talks through the hours we spent going to collect data at farmers' fields. Nothing could pay for it. In addition, I am grateful for his patience and guidance through the development of this project.

I also thank to my committee members Dr. Dustin Harrell, Dr. Thanos Gentimis, Dr. Albert Orgeron, Dr. Richard Johnson, and Dr. Carlos Antonio da Silva Junior. Thank you for the advices, and tips during the process of getting this project done. I hope someday we can work together again!

I also would like to thank LSU AgCenter, USDA and Louisiana Rice Research Board for the funds and help with the development of this project. In addition, I would like to thank some LSU AgCenter employees that somehow supported me during my master's program, Chris Roider, Sandra Stevenson, and Jimmy Flanagan.

I am also thankful for all my friends and colleagues from the LSU Precision Ag Lab and AgCenter Employees, Murilo Martins, Phillip Lanza, Fagner Rontani, Issa Dias, Alex Doylet. In addition, I also would like to thank all the friends I made in my stay here in Baton Rouge, LA. People who made my life much easier and fun, Justin and Carli David, Tyler and Carly Woodard, Johnathan and Brittany Welch, Madison Mills, Jack and Carrie Rigdon, Pastor Bradley and family, Landon, Aaron, Ernesto Ticiano, Cesar and Juliana Menk.

Also, I thank my friends Rodrigo Trevisan and André Colaço who always helped me with my development as a researcher and always were there when I needed. They were my research tutors when I was an undergraduate student in Brazil and will continue to be forever.

Thanks to everyone else that is not listed here but helped me in any way to be here today!

Table of Contents

Acknowledgments iii

List of Tables vi

List of Figures vii

Abstract ix

Chapter 1. Introduction 1
1.1. Precision Agriculture scenario in Rice 6
1.2. Precision Agriculture scenario in Sugarcane 8

Chapter 2. Yield Map Prediction Framework Based on Multispectral Satellite Imagery 10
2.1. Introduction 10
2.2. Material and Methods 11
2.3. Results 34
2.4. Discussion 42
2.5. Conclusion 45

Chapter 3. A Random Forest Approach to Evaluate Yield Map Prediction Using High-Resolution Remote Sensing Imagery Acquired at Different Growth Stages for Rice and Sugarcane 46
3.1. Introduction 46
3.2. Material and Methods 47
3.3. Results 51
3.4. Discussion 59
3.5. Conclusion 62

Chapter 4. Conclusions 64

References 67

Vita……..... 79

List of Tables

Table 2.1. Vegetation indices used in addition to the raw bands from PlanetScope2 for yield map prediction 22

Table 2.2. Cumulative Growth Degree Days (CGDD) and equivalent crop stage selected for rice imagery organization for yield map prediction. 25

Table 2.3. Search-space used as input for *Optuna* hyperparameters optimization – parameters were selected according to *scikit-learn RandomForestRegressor* parameters description 30

Table 2.4. Descriptive statistics for raw and filtered yield for all the study sites 34

Table 2.5. Best performing Random Forest Regression hyperparameters obtained from optimization using the package *Optuna* 37

Table 2.6. Descriptive statistics for the interpolated and predicted yield values for rice test site RC-3 39

Table 2.7. Descriptive statistics for the interpolated and predicted yield values for sugarcane test site SC-3 41

Table 3.1. Rice different prediction scenarios used for the evaluation of the in-season yield map prediction. Each Cumulative Growth Degree Day represents a new image added to the dataset to perform the prediction 49

Table 3.2. Sugarcane different prediction scenarios used for the evaluation of the in-season yield map prediction. Each Cumulative Growth Degree Day represents a new image added to the dataset to perform the prediction 49

Table 3.3. Scott-Knott grouping of means for Person's correlation coefficient and RMSE for each sugarcane CGDD scenario 10-fold Cross Validation 52

Table 3.4. Descriptive statistics for the sugarcane test site (SC2) maps at the different CGDD scenarios 54

Table 3.5. Scott-Knott grouping of means for Person's correlation coefficient and RMSE for each rice CGDD scenario 10-fold Cross Validation 55

Table 3.6. Descriptive statistics for the rice test site (RC3) maps at the different CGDD scenarios 58

List of Figures

Figure 1.1. Adoption of Precision Agriculture activities or tools in US rice according to USDA 6

Figure 2.1. Soil spatial distribution for the three sugarcane fields used for testing the framework proposed in this chapter 12

Figure 2.2. Soil spatial distribution for the three rice fields used for testing the framework proposed in this chapter 13

Figure 2.3. Semi variogram modelling process 17

Figure 2.4. Flowchart for yield map prediction based on satellite imagery and machine learning (ML). Symbols used are in accordance to the BPMN symbology (Chinosi & Trombetta, 2012). GDD – growth degree days; TOA – Top of Atmosphere 18

Figure 2.5. Example of resampling the multiple rasters from different Cumulative Growth Degree Days (CGDD) to a single raster and harvester yield overlaying process. 26

Figure 2.6. Random forest structure and running procedure to perform predictions – n is a maximum number of trees set by the user 28

Figure 2.7. Example of a 4-fold group Cross Validation (CV) that is presented on *scikit-learn* manual (Pedregosa et al., 2012),; whilst a 10-fold group CV was used for the analysis of the yield map predictions. 32

Figure 2.8. Rice training dataset clusters obtained from a *K*-means classification based on Latitude and Longitude. Clusters were used on the 10-fold Group Cross Validation 33

Figure 2.9. Comparison between the yield before and after the filtering process for fields RC2 and SC3 35

Figure 2.10. Number of downloaded satellite imagery for every month and selected CGDDs (°C day) through the sugarcane season Field SC3 36

Figure 2.11. Number of downloaded satellite imagery for every month and selected CGDDs (°C day) through the rice season Field RC2 36

Figure 2.12. Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) for the 10-fold cross validation performed for the rice dataset during training. 38

Figure 2.13. Interpolated (left) and Predicted (right) yield maps for rice in field RC3 39

Figure 2.14. Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) for the 10-fold cross validation performed for the sugarcane dataset during training. .. 40

Figure 2.15. Interpolated (left) and Predicted (right) yield maps for sugarcane in field SC3 41

Figure 3.1. Soil characterization for the sugarcane fields located in Louisiana used for the evaluation of in season yield map prediction 47

Figure 3.2. Soil characterization for the rice fields located in Louisiana used for the evaluation of in season yield map prediction 48

Figure 3.3. Yield map prediction model and evaluation flowchart 50

Figure 3.4. Sugarcane test field (SC2) interpolated and predicted normalized yield maps for the different CGDD scenarios 53

Figure 3.5 Rice test field (RC3) interpolated and predicted normalized yield maps for the different CGDD scenarios 56

Figure 3.6. Two class *k*-means classification for interpolated and Sugarcane 6 and 12 scenarios predicted yield maps. 60

Abstract

Food and energy security are two main topics when it comes to the on-growing world population. Rice and sugarcane play an important role in this scenario since sugarcane can be used for energy production and rice is one of major staple cereals. In this scenario, Precision Agriculture (PA) management strategies aims to improve productivity, efficiency, profitability, and sustainability, and can help agriculture to fulfill the needs of the growing population in a sustainable way. However, yield maps are essential for PA, but its adoption is still very low. Thus, the main objective of this study was to evaluate the potential of satellite imagery and machine learning to predict yield maps that could support the adoption of precision agriculture practices for rice and sugarcane. Consequently, a framework for the data processing, imagery acquisition and machine learning model generation, was proposed and tested. The results presented a high potential for the usage of those techniques, generating yield maps very similar to the ones obtained from yield monitors (RMSE for rice of 0.9Mg.ha^{-1} and for sugarcane 3.14Mg.ha^{-1}). Also, in-season yield map prediction was evaluated for rice and sugarcane. Therefore, the prediction was performed for different growth stages by stacking all the images until a specific date. Sugarcane maps were obtained with a satisfactory accuracy early in the season (May-June) (no statistical significance when compared to the predicted maps of the end of the season) whilst for rice the yield maps with the lowest errors were only obtained late in the season. Therefore, sugarcane maps obtained early in the season could be used for in-season management of the crop. On the other hand, the in-season applicability for rice yield maps were limited since accurate maps were obtained at late ripening. However, this information could still be used for harvest planning and nitrogen application on the second harvest of Louisiana's rice. In general, the framework proposed presented a high potential to be used for yield maps

prediction. Furthermore, yield maps, an important tool for PA, were obtained with low errors RMSE of 0.83 and 3.14 $Mg.ha^{-1}$ for rice and sugarcane, respectively.

Chapter 1. Introduction

Rice and sugarcane are historically and economically important crops for the state of Louisiana. The first time rice production in Louisiana was mentioned in the literature was in 1710, whilst in 1795 sugar was first manufactured in Louisiana (Lee, 1960; Prichard, 1939). Therefore, both crops are attached to Louisiana's history and development.

In 2018, sugarcane was produced in 24 parishes in South Louisiana, totaling over 182,000 hectares (Deliberto et al., 2019). In the same year, approximately 175,000 hectares of rice were produced in Louisiana (Louisiana Rice Research Board & LSU AgCenter, 2018). The combined economic impact of rice and sugarcane production in 2018 was around $1.5 billion (LSU AgCenter, 2018), demonstrating the importance of these commodities to Louisiana's agriculture.

The world's food and energy security can be affected by the production and use of both crops. Rice is considered one of the major staple cereals in the world and about 3.5 billion people around the world have 20% of their daily caloric intake based on its consumption (IRRI, AfricaRice, & CIAT, 2010). Sugarcane is primarily grown as a source of sugar (Louisiana scenario); however, this crop has the potential to produce biofuel, and has been utilized by different countries, especially Brazil (Khan & Khan, 2019).

Based on the importance of both crops and considering the projection that the world population will reach 9.1 billion by 2050 (FAO, 2009), improvements to food and energy production will be necessary to fulfill the needs of the population in the future. In this scenario, the adoption of a management strategy known as Precision Agriculture (PA) can be used to meet these needs. According to the International Society for Precision Agriculture (ISPA, 2019) PA is defined as:

> "…a management strategy that gathers, processes and analyzes temporal, spatial and individual data and combines it with other information to support management decisions according to estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production."

Based on this definition, the objectives of PA match the current and future changes that agriculture must undergo to meet production and environmental goals. PA is based on the analysis of temporal and spatial variability, which translates to site-specific crop management, a strategy known for a long time. In 1929, a circular published by the University of Illinois proposed a methodology for the development of a grid-based soil data collection, analysis and lime zone application based on the spatial variability of the soil acidity (Linsley & Bauer, 1929).

When the methodology proposed by Linsley & Bauer (1929) was implemented, Global Navigation Satellite Systems (GNSS) were not available; therefore, the tracks to be followed for soil sampling and the final lime application map were generated using measurements and cardinal directions. With the development and release of civilian GNSS, Linsley & Bauer's (1929) methodology for soil collection was adapted to use the GNSS receptors.

Simultaneously, and due to the development of GNSS, other technologies that support PA were developed. One example is the yield monitor sensor within a modern grain combine, which measures yield as the crop is being harvested, and combined with GNSS receptors, records the geographical location for each yield value measured (Vega et al., 2019). Using this type of sensor, the crop itself is used as a bioindicator of spatial variability, and is one of the most common and important tools used for the development of PA, the yield map (Adamchuk et al., 2011; Vega et al., 2019).

In addition to yield monitors, the usage and development of other sensors are also especially important for PA. Sensing of agricultural crops can be divided into two types, remote sensing, and proximal sensing. According to Adamchuk et al. (2011), remote sensing involves

the usage of sensors installed on airborne or satellite platforms; while proximal sensing requires sensors to be used close to the object of study, or even in contact to it.

Regarding the usage of remote and proximal sensing, optical sensors for plant growth and health studies, have been a primary focus. A great deal of research has focused on the development and evaluation of different vegetation indices to estimate plant biophysical characteristics. One of the most known and commonly used indices is the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973) which uses the spectral bands red and near-infrared to generate the index. A limitation of NDVI is a phenomenon called saturation, which translates to its non-linear relationship with some plant characteristics such as Leaf Area Index (LAI) and aboveground biomass (Huete et al., 2002). This phenomenon usually happens when crop density is moderate to high causing a limitation of NDVI usage in later stages of the crop. Therefore, research has focused on the generation of vegetation indices which minimize the effect of saturation observed for NDVI. One such index is the Wide Dynamic Range Vegetation Index (WDRVI) which utilizes the same spectral bands as the NDVI, but includes a coefficient in the NDVI equation intended to raise the sensitivity of the index at high and moderate crop densities (Gitelson, 2004). Other indices were developed to reduce the background effect of soil, such as the Optimized Soil Adjusted Vegetation Index, which adds a coefficient to the NDVI equation with the objective to reduce the soil effect on the NDVI (Rondeaux, Steven, & Baret, 1996). These modified indices allow more precise mapping of spatial variability within the crop in the early (more affected by the background – soil or water) and late stages (usually related to high crop density) of production.

The use of remote and proximal sensing in PA and applied research, such as the development of vegetation indices, supports the study of spatio-temporal variability of the crop within the

growing season, as opposed to obtaining information after harvesting, or having to perform time consuming and costly sampling procedures within the crop season. In this scenario, PA strategies have been developed based on proximal sensing for in-season biomass estimation and nitrogen application (Holland & Schepers, 2010; Raper, Varco, & Hubbard, 2013; Solie et al., 2012), soil chemistry and physics evaluation (Molin & Tavares, 2019; Serrano et al., 2014; Tavares et al., 2020), weed control (Peteinatos et al., 2014; Sui et al., 2008), plant phenotyping (Jiang et al., 2018; Pallottino et al., 2019; Rischbeck et al., 2016) and others.

When it comes to remote sensing, satellite imagery and Unmanned Aerial Vehicles (UAVs) provide an effective tool to conduct remote sensing of crops. The resolution of imagery collected with both satellite solutions and UAVs can vary greatly and depends on several factors. In general, imagery from paid satellite imagery providers have higher resolution products as compared to no cost sources. Planet, a paid satellite service, has multiple imagery products available. One such product, PlanetScope 2 (PS2), provides imagery with a 3.7m pixel size and captures additional images of the site on a daily basis. PS2 imagery contains four different spectral bands: red, green, blue and near infrared. Another product from Planet is Skysat. Skysat imagery provides a higher spatial resolution as compared to PS2 (0.5m vs. 3.7m); however, revisit time only occurs in 4 to 5-day intervals. Skysat imagery provides the same four bands as PS2 (Planet Labs Inc., 2020).

No cost satellite imagery is widely available but results in lower spatial resolution of imagery. Sentinel-2 (S2) constellation provides the same spectral bands obtained from PS2 and Skysat, at a pixel resolution of 10m and revisit frequency of five days. S2 also offers nine additional bands at 20m and 3 bands at 60m resolution (Gatti et al., 2018). Satellite products are very advanced and can provide a lot of useful information for agriculture. One major advantage

of using satellites is the frequent and autonomous revisit frequency, thus providing a new image every few days. Atmospheric conditions can provide a limitation for satellite imagery. Excessive cloud cover negatively impacts the quality of the data, especially for satellites with a lower revisit frequency.

The use of UAVs present an advantage over satellites, especially when atmospheric conditions are unfavorable, since the flight time and date are determined by the operator who can choose the best day to obtain the imagery for the work site. Imagery resolution obtained with the multispectral sensors embedded on the UAV is very high (e.g.: UAV imagery can be acquired at a pixel size of 3cm). However, one limitation of utilizing UAVs to conduct remote sensing is the need for an operator (pilot) to launch and observe the aircraft during the flight, even though the actual flight can be autonomous.

Proximal and remote sensing studies for PA evaluate the correlation between the information gained from the sensors with crop biomass and yield. These studies in conjunction with current higher computational capabilities, have provided researchers with tools that have been utilized to forecast and predict yield in wheat (Nevavuori et al., 2019; Peerlinck et al. 2018; Zhao et al., 2020), corn (Khanalet al., 2018; Kim et al., 2020; Martin et al., 2012; Peralta et al., 2016; Shiratsuchi et al., 2011) and soybean (E. E. da Silva et al., 2020; Maimaitijiang et al., 2020; Sun et al., 2019). These studies utilize different techniques to derive yield prediction (classical machine learning, deep learning, multivariate regression, and others) and different sensors (high- and low-resolution satellite imagery, UAV multispectral images, active canopy sensors, and others) for the prediction. Situationally, weather data or crop phenology were included as inputs to improve the predictions. Moreover, predictions at different scales were also performed, including within-field yield variation – yield maps.

As sensors have evolved in complexity, the amount of data captured has increased exponentially. Consequently, to analyze this massive amount of data researchers rely on techniques such as Artificial Intelligence (AI) to aid in data analysis. Agriculture specific, AI methods and procedures have been developed and will aid in the achievement of PA objectives.

1.1. Precision Agriculture scenario in Rice

According to the United States Department of Agriculture (USDA) Economic Research Service report for crop practices (USDA, 2019), yield monitors were adopted by approximately 58% of domestic rice producers in 2013; however, the utilization of data collected to generate yield maps was not measured. In a similar report on 2006 harvesting practices, the USDA Economic Research Service (2019) reported 30% of rice hectarage was harvested using combines equipped with yield monitors, but only 9% of those hectares had yield maps created. An increase in the usage of variable rate technology (VRT), autosteering system and georeferenced soil sampling was observed from 2006 to 2013 according to the USDA report (Figure 1.1).

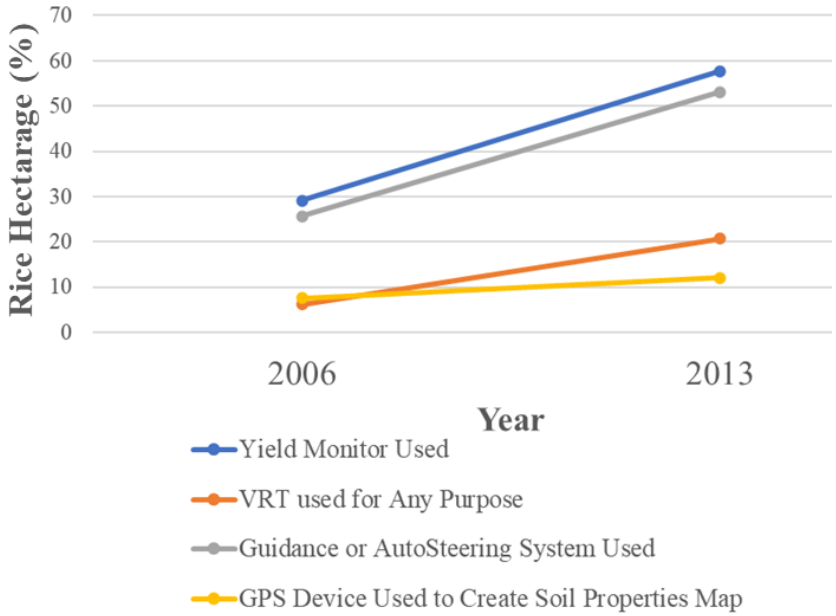

Figure 1.1. Adoption of Precision Agriculture activities or tools in US rice according to USDA

A similar scenario to the one in the United States (US), is also reported by Higgins et al. (2017) in Australia. The adoption of PA is scarce in Asia (Lowenberg-DeBoer & Erickson, 2019); however, many studies developed in China can be found proposing methods for in-season nitrogen application for rice, based on the usage of remote and proximal sensing (e.g. Yao et al., 2012; G. Zhao et al., 2013; Q. Zhao et al., 2015).

In general, by analyzing rice production areas around the world, it is possible to conclude that, efforts have been made to promote the adoption of PA tools for rice production, yet the majority of PA tools are considerably underutilized. Despite the availability of combine sensors that can provide yield maps, farmers are more likely to use other tools such as variable rate technology since those technologies are more ready to use than the yield assess from the combines (Griffin et al., 2019).

Most studies evaluating the potential of yield mapping in rice have focused on the use of multispectral sensors embedded in UAVs (Wan et al., 2020; F. Wang et al., 2019; Zhou et al., 2017), usage of satellite imagery (Aboelghar et al., 2011; Wang et al., 2012; Q. Zhao et al., 2015), and the use of active canopy sensors (Harrell et al., 2011; Kanke et al., 2016).

Even though the importance of yield maps might be neglected, they are essential for the establishment of site-specific management programs because of their potential to address the spatio-temporal variability within fields (Maestrini & Basso, 2018a, 2018b). Consequently, considering the PA scenario in rice, the importance of yield maps and potential of remote and proximal sensing for yield prediction requires the development and evaluation of methodologies for rice map yield prediction is necessary to meet the needs of a growing world population.

1.2. Precision Agriculture scenario in Sugarcane

PA adoption for sugarcane is not as well noted as it is for other crops such as rice. However, through the studies being developed for this crop it is possible to delineate the PA adoption for this crop.

Sugarcane is a perennial crop, and the crop is harvested several times from a single planting event. In Louisiana, three to four ratoon crops are typically harvested prior to the field being fallowed, then replanted (Gravois et al., 2016). Unlike grain crops, where only the grain is harvested, the entire plant is harvested and is processed to make sugar. The first sugarcane yield monitor prototype was released in 1997 (Corredo et al., 2020) while yield monitors for grain crops have been commercially available since the early 1990's (Fulton et al., 2009). Even though sugarcane yield monitors have been developed since 1997, their availability and usage is scarce (Corredo et al., 2020).

Yield maps are essential to the establishment of PA programs, consequently the low availability of yield monitors for sugarcane has delayed the development of several PA tools for this crop. Recently, new John Deere® sugarcane combines have the option to add yield monitors (Darr et al., 2015); however, the development and use of sugarcane yield maps are rare.

The lack of available sugarcane yield maps has not prevented the development of some PA tools for the crop. PA tools such as auto guidance, VRT, proximal and remote sensing as potential technologies to improve sugarcane management strategies and yield are being utilized sparsely. A survey from Brazil's main sugar-ethanol industry in São Paulo state reported that among PA technologies, the usage of satellite imagery had the highest adoption rate, followed by auto guidance for harvesters; furthermore, georeferenced soil sampling and VRT also showed high adoption rates (Silva et al., 2011). Passalaqua & Molin (2020) showed the importance of

auto guidance technologies to reduce ratoon damage caused when transshipment trailers are driven over adjacent rows in Brazil.

Other research has evaluated the potential of using VRT in sugarcane for lime application (Johnson & Richard, 2010), nitrogen application (Amaral et al., 2015; Amaral et al., 2018; Lofton & Tubaña, 2016; Sofonia et al., 2019), weed control (Tangwongkit, 2006) and others. Additionally, studies have been conducted to evaluate the potential of remote and proximal sensors for yield prediction, specifically active canopy sensors (Amaral et al., 2014; Lofton et al., 2012; Portz et al., 2012), satellite imagery (Bégué et al., 2010; Fernandes et al., 2011; Rahman & Robson, 2020), UAV multispectral imagery, and LiDAR (Sanches et al., 2018; Sofonia et al., 2019).

Molijn et al. (2019) published a commercial field scale study for within-field yield prediction and using optical and radar satellite imagery. They showed a high potential for the prediction of yield zones inside sugarcane fields. However, the biggest challenge addressed by this study was the assessment of yield, performed through sampling. Currently, this challenge can be overcome through the usage of commercially available sugarcane yield monitors.

According to the PA scenarios presented for rice and sugarcane, but also based on the importance of yield maps for this management strategy, the development of research regarding the prediction of yield maps for both crops is necessary. In order to show value to rice and sugarcane farmers, yield maps need to be easily accessed and not require many inputs or actions from them. These tools need to be autonomously collected, as well as the scalability of the prediction system must meet the farmer's need. Thus, the main objective of the present study is to evaluate the potential of high-resolution satellite imagery and machine learning in the prediction of yield maps for sugarcane and rice focused on PA applications.

Chapter 2. Yield Map Prediction Framework Based on Multispectral Satellite Imagery

2.1. Introduction

The green revolution allowed the increase of agricultural production worldwide based on improvements in plant breeding, usage of agricultural inputs and mechanization (Frankel, 1971). However, through the indiscriminate and unscientific usage of the technologies brought by this revolution, negative effects in the production system arose (Bhakta, Phadikar, & Majumder, 2019), namely reductions of soil health, fertility, increased erosion, and environmental pollution caused by the leaching of agricultural inputs (Herder et al., 2010).

Over the years consumers have increased pressure over the farming sector, and have demanded healthier, environmentally friendly and higher quality products (Dimitri et al., 2005). Consequently, agricultural production systems have been pushed to fulfill the market demands. In this scenario, PA has an important role, since its main purposes are to improve agriculture efficiency, productivity, and sustainability through the study of spatial and temporal variability while increasing profitability (Shannon et al., 2018).

Yield maps are one of the most important tools for the understanding and managing the spatial and temporal variability within a field (Blackmore et al., 2003; Cerri & Magalhães, 2005; Momin et al., 2018) however; it is among the most underutilized tools in PA (Lowenberg-DeBoer & Erickson, 2019). In the United States, half of the owners of combines equipped with yield monitors usually do not generate yield maps (Lowenberg-DeBoer & Erickson, 2019). The low usage of such an important tool might be related to how readily this information is to the farmer compared to other tools such as VRT based on soil fertility or other grid sampling-based products (Griffin et al., 2019). Another factor is the lack of yield monitors developed for some crop or scarce usage in them.

There is a potential for the use and development of yield map prediction systems that can offer ready-to-use information and yield maps in crops where yield monitoring is still a challenge. Yield prediction has been explored in different cropping systems and remote sensing presents a promising mechanism for this purpose (Ali et al., 2019; Azzari et al., 2017; Bégué et al., 2010; Haghverdi et al., 2018; Hamada et al., 2015; Peralta et al., 2016; Son et al., 2013).

Considering that rice and sugarcane are two crops economically important for Louisiana, and that a limited number of studies have explored the potential of intra field satellite-based yield prediction for those crops, they were selected for this study. Therefore, the objective of this study was to evaluate a framework for the development of a yield map prediction model based on satellite high-resolution resolution imagery and machine learning for rice and sugarcane.

2.2. Material and Methods

2.2.1. Study Sites

To evaluate the framework proposed, a total of six fields were used as inputs. Three of those fields were from sugarcane (SC1 [3.9 ha], SC2 [5.5 ha] and SC3 [5.2 ha]) and three from rice (RC1 [29.3 ha], RC2 [41.7 ha], RC3 [17.1 ha]).

- Sugarcane Study Site Description

The Enterprise Sugarcane Factory located in Jeanerette, Louisiana, (N 29°53’30.57” W 91°43’54.5”), provided the sugarcane data. Crop stage for the three fields was third ratoon. According to the National Centers for Environmental Information (NCEI) monthly rainfall averaged 133 mm and yearly rainfall averaged1543 mm for the region. While, the average minimum temperature is 14.7 °C and average maximum temperature is 25.6 °C. It is important to mention that even though the minimum average temperature for the region is not below 0 °C, during winter season the minimum temperature can reach negative values, causing damage to sugarcane plants (Eggleston et al., 2004).

The soil type for two of the fields was classified as a Jeanerette silt loam, whereas the third field was classified as a Patoutville silt loam according to the Soil Survey Geographic Database (SSS, NCRS, & USDA, 2020) (Figure 2.1).

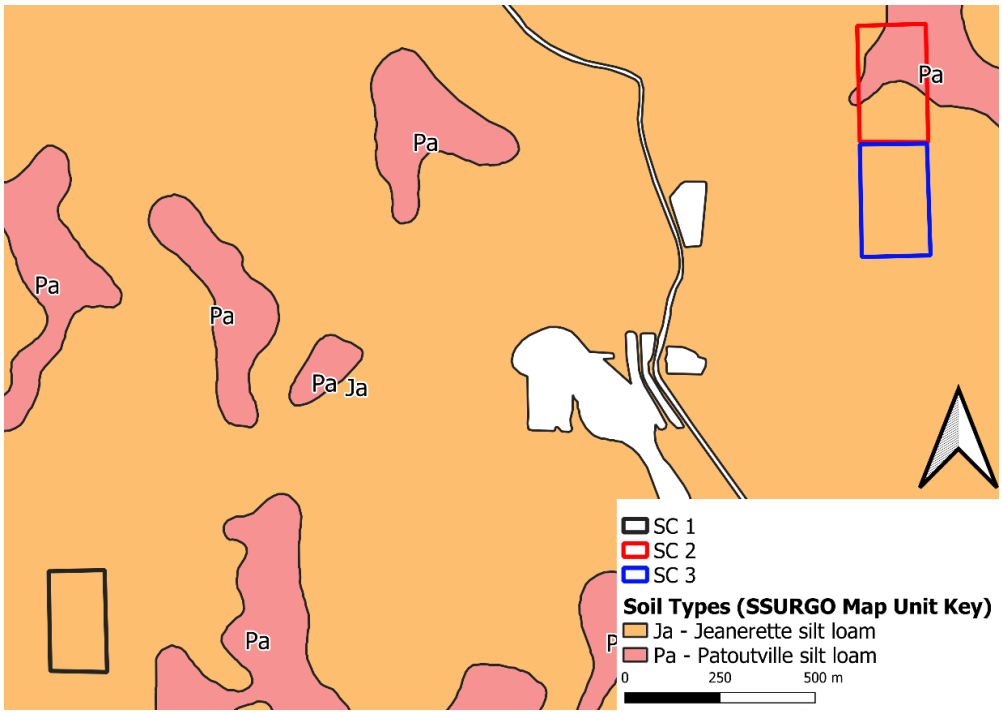

Figure 2.1. Soil spatial distribution for the three sugarcane fields used for testing the framework proposed in this chapter

- Rice Study Site Description

A farm located in Gueydan, Louisiana, (N 30°02’57.0” W 92°33’20.9”), provided the rice data. According to the NCEI yearly rainfall averaged 1591 mm, and minimum and maximum temperatures averaged 15.6°C and 25.8°C, respectively. However, within rice production season (March – August) the rainfall averaged approximately 100 mm (March) and 175 mm (June), respectively. Whilst in the beginning of the season temperatures can be as low as 11°C, which can cause damage to the seed and crop affecting its development (Krishnan, et al, 2011; Yang et al., 2019). Conversely, temperatures higher than 30°C can occur in the end of the

season, which can also impact the yield and grain quality, especially if those high temperatures occur at night (Krishnan et al., 2011).

The three rice fields used for the analysis of the proposed framework are located within the same region. Soil type differed for each field (Figure 2.2). However, the texture for all soils were classified as silt loam, with drainage classification from somewhat poorly drained to poorly drained (SSS et al., 2020). The poor drainage of those soils is a very important characteristic for the water management of rice production in this region of Louisiana, since the crop is produced in a flooded environment.

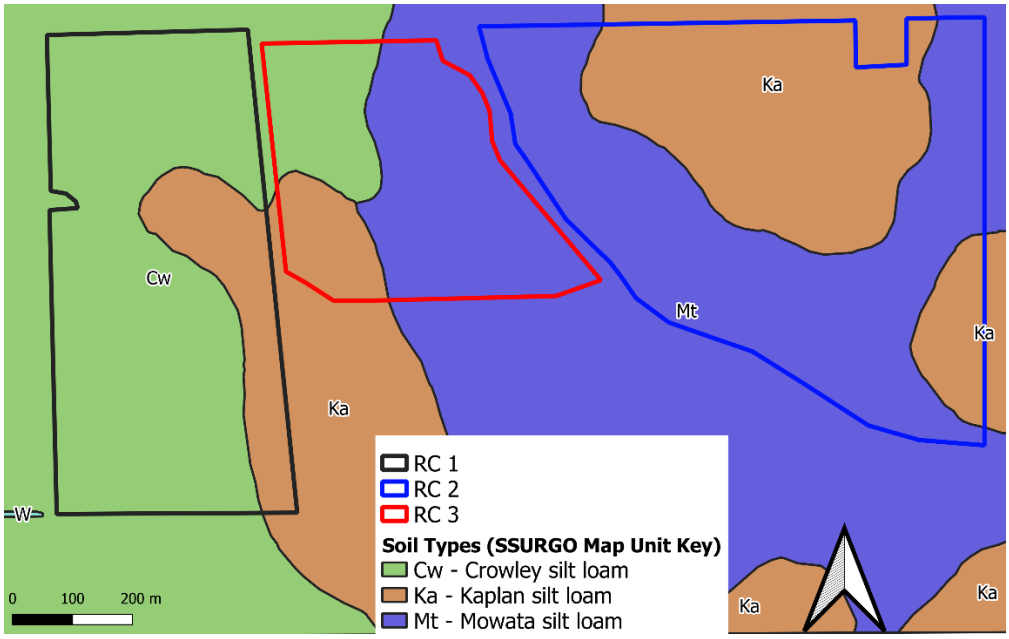

Figure 2.2. Soil spatial distribution for the three rice fields used for testing the framework proposed in this chapter

2.2.2. Prediction System and Framework Design

The framework proposed had the purpose of performing yield map prediction from using the minimal user input, allowing a complete automatization of the prediction process. Based on these premises, Figure 2.4 presents a flowchart, built from the symbology used by BPMN

methodology (Chinosi & Trombetta, 2012), that explains the framework proposed and tested in this chapter.

- Yield Data and Processing
    - Yield from Harvester

In precision agriculture, yield maps are usually obtained from harvesters that are equipped with yield monitors systems. Commonly, the data exported by those systems are points. Each point will have the recorded flow acquired by the yield sensor, machine speed, machine swath, calculated yield, plus other information about machine status and harvesting process. In the framework proposed, the yield obtained by those monitors was considered as benchmark (target variable) to build the yield map prediction model.

The yield obtained for rice and sugarcane were georeferenced points which only the geospatial information and the calculated yield were used. According to the farmers that provided the data, the yield monitors for both crops were calibrated before the harvesting process. Yield data from both crops were downloaded from commercial database platforms used by the farmers. Thus, the format of the raw yield data used in this study was a shapefile, a common format used for geospatial data (ESRI, 1998). Yield data had approximately a density of 3,000 points per hectare (average distance between passes of 1.4m and between consecutive points of 1.5m) and 1500 points per hectare (average distance between passes of 10m and between consecutive points of approximately 1m) for sugarcane and rice, respectively, the data collection rate for both crops was 1Hz.

    - Yield Global and Local Filtering

Errors related to yield monitoring systems are well known and have been studied for a long time by several authors (Blackmore & Moore, 1999; Lyle, Bryan, & Ostendorf, 2014;

Maldaner, Molin, & Canata, 2016). In order to remove those errors, a filtering process needs to be performed. Also, to remove the errors listed by previous literature, the need of global and local filtering have been mentioned by a few authors (Leroux et al., 2018; Lyle et al., 2014; Maldaner et al., 2018; Menegatti & Molin, 2004; Sudduth & Drummond, 2007; Vega et al., 2019).

Therefore, for the dataset used in this chapter, a global and local filtering system was built under the language R (R Core Team, 2019). The system developed performs a global filtering using the interquartile range (IQR), thus any values lower than first quartile minus 1.5 times IQR or higher than third quartile plus 1.5 times IQR are considered outliers (Tukey, 1977). In addition, by setting a searching radius (Spekken, Anselmi, & Molin, 2013), a local isotropic filtering took place by calculating low and upper limits based on the local median and a set variation limit (Maldaner et al., 2018). A development of a customized function within the programming language R took place to perform the local filtering. Using the packages *sf* (Pebesma, 2018) and *future* (Bengtsson, 2020) a moving radius of search was developed in which the yield value for the point in the center of the radius was compared to the overall median of the points within the area of search. If the values from center point was outside the local median limits, it was considered an outlier and removed (Maldaner et al., 2018).

The radius and maximum variation limit values used for the dataset were set according to an analysis of the histogram data distribution, mean, median, coefficient of variation, standard deviation and final maps obtained from the filtering process. Different values were tested until filtering results similar to the presented in the literature were obtained (Maldaner et al., 2016; Spekken et al., 2013; Sudduth et al., 2012). A radius of 20m was used for both crops whilst for maximum variation limit values of 10% and 15% were set for sugarcane and rice, respectively.

- Filtered Yield Interpolation

Since the yield data obtained from the harvesters are points, to obtain yield maps interpolation is commonly used in PA (Blackmore & Moore, 1999). Within the existent interpolation methods, kriging is considered the most accurate estimator for interpolation (Isaaks & Srivastava, 1989); therefore it was used in this chapter.

As presented on Figure 2.4, in the proposed framework, the interpolated values are not used during the model generation, but on the evaluation of the results obtained from the prediction model. During this step, the predicted and the interpolated yield maps are compared against each other.

To perform the interpolation of the filtered yield data, the use of the R package *gstat* took place (Gräler, Pebesma, & Heuvelink, 2016). Using this package, a semi variogram for the dataset can be obtained. A semi variogram is a graph which represents the semi variances of a specific parameter in function of the distance between the neighboring points' values obtaining a cloud of semi variances (Figure 2.3a). Then this cloud is averaged by setting a bin width (also known as "lag"). The graph obtained from this averaging process is known as experimental or sample variogram (Figure 2.3b), which is used to adjust a mathematical model that will represent the spatial dependence of a specific parameter (Figure 2.3c) (Hengl, 2009).

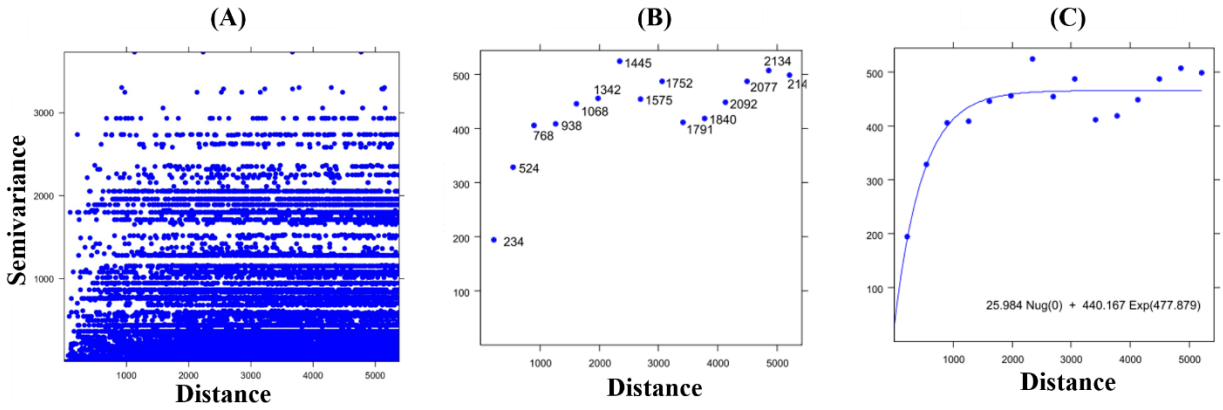

Figure 2.3. Semi variogram modelling process. A- cloud of pair of points presenting the semi variance; B – Experimental variogram (pair of points aggregated to lags of 300m); C – adjustment of variogram model to the points. Adapted from Hengl (2007)

In this study, the filtered yield from the harvesters were used as the parameter which semi variance was calculated. Three different variogram models, exponential, spherical and gaussian, were tested and the best one was used for the interpolation. The most appropriate model was obtained based on the analysis of the sum of squared errors (SSErr) of the fitted model. Based on the analysis of the SSErr for the dataset, a spherical model presented to be the most appropriate semi variograms model for the yield data from both crops and was used for the interpolation and obtainment of the yield maps.

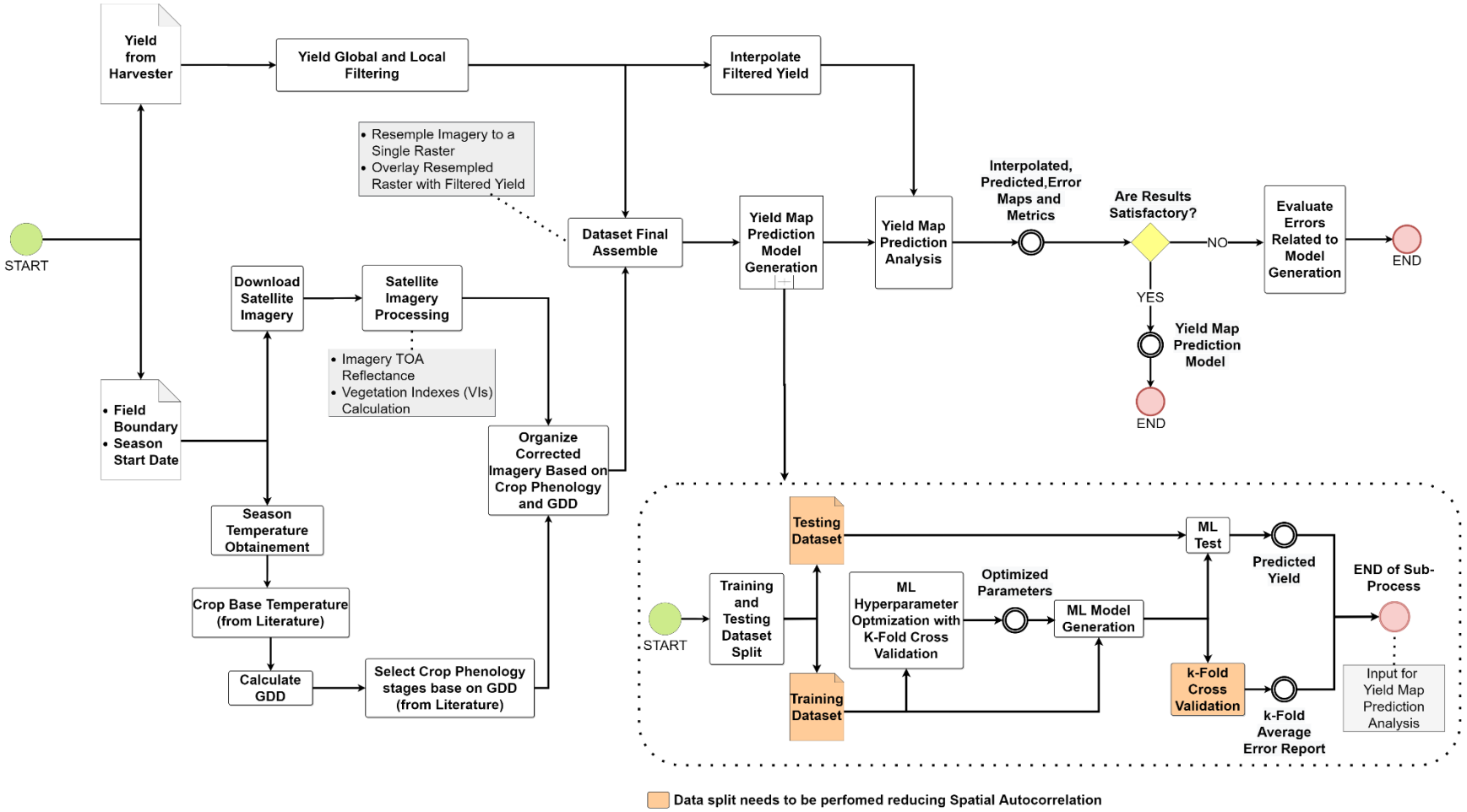

Figure 2.4. Flowchart for yield map prediction based on satellite imagery and machine learning (ML). Symbols used are in accordance to the BPMN symbology (Chinosi & Trombetta, 2012). GDD – growth degree days; TOA – Top of Atmosphere

- Imagery Download and Processing

Based on the field boundary and season starting date provided, the imagery for the sites where yield data was assessed were downloaded and processed.

  - Satellite Imagery Download

There are multiple satellite options from which data can be acquired and processed. Since the focus of the proposed framework is to predict yield maps focusing on PA usage, it is important to consider that higher spatial resolution imagery will result in a higher resolution in the final yield predicted map. Also, to guarantee that good images will be available through important crop stages during the season, more frequent revisits of the satellite over the studied site is desirable.

For the evaluation of the framework, PS2 came in use, because it provides daily revisit frequency and a pixel size of approximately 3.7m. The spectral bands acquired by the sensors embedded at PS2 are blue (455-515 nm), green (500-590 nm), red (590-670 nm) and near infrared (NIR; 780 – 860 nm).

The download of satellite imagery itself can be done using the platforms provided by the organizations responsible for the systems. However, this can be a time-consuming process, especially when the number of images is high. Fortunately, most available satellites databases can be accessed by an Application Programming Interface (API), an easier and faster way to download larger numbers of images.

The PS2 data used was downloaded using the API provided by Planet Team (2018). A customized script programmed under the language Python 3.0 (Van Rossum & Drake, 2011) and the python package *planet* (Planet Team, 2018) was used to access Planet database. Images with

no more than 10% of cloud coverage were downloaded from each field specific location and for the entire crop season.

- Satellite Imagery Processing

The satellite assets downloaded using the API were converted to top of atmosphere (TOA) reflectance following the guidelines provided by Planet Labs (Planet Team, 2017). This correction was performed using a customized script under R language. The conversion is based on the coefficient values provided for each asset band; information available in the metadata downloaded with the imagery. Those coefficients were calculated based on the atmospheric and aerosol models from MODIS 6SV (Kotchenova et al., 2006)

Also, based on previous yield map prediction studies (E. E. da Silva et al., 2020; Maimaitijiang, Sagan, Sidike, Hartling, et al., 2020; Wan et al., 2020) and on the spectral bands available on the imagery used, a list of vegetation indices (VIs) was calculated for each image (Table 2.1). The same customized script used for the TOA conversion also calculated the different vegetation indices for each asset downloaded.

A total of 8 VIs were calculated. The Normalized Vegetation Index is one of the most wide used VI; therefore, was included in this study. However, this index is known for a phenomenon called saturation, which causes a nonlinear relationship between biomass and leaf area index (LAI) at high biomass. One of the causes of this phenomenon is addressed in the literature to the chlorophyl absorption of the red waveband. Therefore, to overcome this issue a few other indices were used in the present study: Green Chlorophyll Index, Green Normalized Difference Vegetation Index, Simple Ratio, Wide Dynamic Range Vegetation Index and Enhanced Vegetation Index. The usage of those indices is important since they increase the map accuracy late in the crop season. However, in the beginning of the crop development, the

background – such as soil and water – can affect the performance of the vegetation mapping. Therefore, Optimized Soil Adjusted Vegetation Index and Modified Soil Adjusted Vegetation Index 2 were calculated at each image to reduce the soil effect on the vegetation mapping. Unfortunately, the sensors embedded on the satellite used in this study did not provide bands such as Short-Wave Infrared (SWIR) that could be used to reduce the water background effect for rice.

After the correction and calculation of the VIs listed on Table 2.1, a total of 12 variables were obtained for each downloaded asset (red, green, blue, and near-infrared bands plus the 8 VIs). All those variables were used as inputs for the yield map prediction.

- Growth Degree Days (GDD) Calculation and Crop Phenology Stages

In the proposed framework, the imagery organization uses the crop thermal time, expressed in cumulative growth degree days (CGDD). This information is relevant because according to Trudgill et al. (2005), thermal time provides the possibility of comparing the development of organisms, such as plants, in terms of its development within the temperature range of base temperature and optimum temperature. Therefore, thermal time is commonly used because of its relation with the crop phenology stages, since the thermal units to reach a specific ontogenetic stage are fairly constant (Purcell, 2003).

Table 2.1. Vegetation indices used in addition to the raw bands from PlanetScope2 for yield map prediction

Vegetation Index	Nomenclature	Equation	Reference
EVI	Enhanced Vegetation Index	$2.5 * \frac{NIR - Red}{(NIR + 6 * Red - 7.5 * Blue + 1)}$	Huete et al. (2002)
GCI	Green Chlorophyll Index	$\frac{NIR}{Green} - 1$	Gitelson, Gritz, & Merzlyak (2003)
GNDVI	Green Normalized Difference Vegetation Index	$\frac{(NIR - Green)}{(NIR + Green)}$	Gitelson & Merzlyak (1998)
MSAVI2	Modified Soil Adjusted Vegetation Index 2	$\frac{2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2 - 8 * (NIR - Red)}}{2}$	Qi, Chehbouni, Huete, Kerr, & Sorooshian (1994)
NDVI	Normalized Difference Vegetation Index	$\frac{(NIR - Red)}{(NIR + Red)}$	Rouse, Hass, Schell, & Deering (1973)
OSAVI	Optimized Soil Adjusted Vegetation Index	$\frac{(NIR - Red)}{(NIR - Red + L^*)}$	Rondeaux, Steven, & Baret (1996)
SR	Simple Ratio	$SR = \frac{NIR}{Red}$	Birth & McVey (1968)
WDRVI	Wide Dynamic Range Vegetation Index	$\frac{(a^{**} * NIR - Red)}{(a * NIR + Red)}$	Gitelson (2004)

*L coefficient used was 0.16 (Rondeaux et al., 1996) ** a coefficient used was 0.2 (Henebry, Viña, & Gitelson, 2004); NIR stands for Near Infra-red

In this scenario, selecting CGDD that are related to specific crop stage or crop development, allows the comparison over time and space of the crop characteristics. Therefore, organizing the imagery based on CGDD facilitates the final yield map prediction model scalability over different fields in different locations and years, while it also carries a phenological information.

To calculate GDD, the acquisition of minimum, maximum, and the crop base temperature is necessary (Mcmaster, Wilhelm, & Wilhelm, 1997). Crop base temperature can usually be found on literature. Whilst environment maximum and minimum temperature can be commonly obtained from ground weather station (GWS). In the United States, where the study sites for this chapter are located, the National Oceanic and Atmospheric Administration (NOAA) provides an API that allows the access and download of their WS database.

To test the framework, maximum and minimum temperatures were downloaded from NOAA database for the different study sites. Besides that, the base temperature used for the GDD calculation for sugarcane was 18°C (Lofton et al., 2012; Teruel, Barbieri, & Ferraro Jr., 1997) and for rice 10°C (Gao, Jin, Huang, & Zhang, 1992). Based on this information and the season start date, CGDDs were calculated.

GWS data were downloaded by accessing NOAA's database using their API through a customized script under the language R. The package *rnoaa* (Edmund, Chamberlain, & Ram, 2014) took place to retrieve all the weather stations within a 100km radius around the study sites. Then, the closest GWSs that had minimum and maximum temperature information available had their data downloaded. Based on method two described by Mcmaster, Wilhelm, & Wilhelm (1997)the CGDD values were calculated based on GWSs located at 20 km and 8 km for the rice and sugarcane fields, respectively.

- Final Dataset Assemble
    - Imagery Organization Based on Crop Phenology/CGDD relation

To organize the imagery based on Phenology and CGDD relation, it was necessary to set an interval of GDD or specify CGDD values that would be used to select the images for the final dataset. As mentioned in Figure 2.4 this CGDD selection should be done using previous literature studies that present the crop phenology stages and thermal time.

For the two crops used in this chapter, two sets of specific CGDD values were chosen to develop two separate models. According to previous research, the sugarcane leaf appearance rate varies from 70°C day through 130°C day during the season (Inman-Bamber, 1994). Therefore, a value of 115°C day, and a total of 14 values were set, ensuring that the latest CGDD would be at least 10 to 15 days before the harvest date. Also, the chosen interval assured that images from important crop development stages were included as inputs for the model, such as the peak of tiller number (500°C day to 800°C day) and its stabilization (around 1200°C day)(Bonnett, 2013; Inman-Bamber, 1994; Keating, Robertson, Muchow, & Huth, 1999; Marin & Jones, 2014).

For the rice dataset, instead of using a GDD interval, specific CGDD were set to match important crop stages and crop development. Therefore, the values chosen are presented at Table 2.3 with corresponding crop development stage. The CGDD were set according to previous research that focused on rice yield prediction and crop development (Harrell, Tubaña, Walker, & Phillips, 2011; Q. Jiang et al., 2019; Wan et al., 2020).

Table 2.2. Cumulative Growth Degree Days (CGDD) and equivalent crop stage selected for rice imagery organization for yield map prediction.

CGDD	Crop Stage
265	3rd leaf
400	5th leaf/1st tiller
660	Jointing
820	Panicle Initiation
990	Panicle Differentiation
1250	50% Heading
1500	Early Ripening
1750	Late Ripening

Therefore, using the CGDD and the dates in which they were achieved, the images were selected for each crop. Thus, a total of 14 and 8 images were used for each field from the sugarcane and rice datasets, respectively. In addition, after selecting the images for each field, they were resampled to a single raster containing all bands and VIs for all the GDDs (Figure 2.5). The resampling process was performed using a customized R script based on the usage of the package *raster* and function *resample* (Hijmans, 2020). The function *resample* can perform the resampling processes for two rasters at each call of the function. Therefore, the first image obtained for the season was used to store the values of the remaining imagery – the *resample* function was called 13 and 7 times for sugarcane and rice, respectively. As a result, a single raster with 168 (14 different CGDD times 12 variables per CGDD) and 96 (8 different CGDD times 12 variables per CGDD) variables were obtained for sugarcane and rice, respectively.

- Binding Imagery and Filtered Yield data

Finally, to obtain the dataset that will be used to generate the yield map prediction model, the filtered yield needs to be added to the single raster. Therefore, based on the spatial location of both, yield and raster, the information was overlayed. Since yield data was collected at a high frequency (1Hz), more than one yield observation can be grouped inside the same pixel. When

this situation is encountered, an average of the yield values related to all the points that are within the same pixel is calculated. Then the resulted value is addressed as the yield for the pixel (Figure 2.5). Since sugarcane harvesting process is performed by row (every 1.4m), there is a higher density of yield points when compared to rice (which machine harvesting platform had approximately 10m).

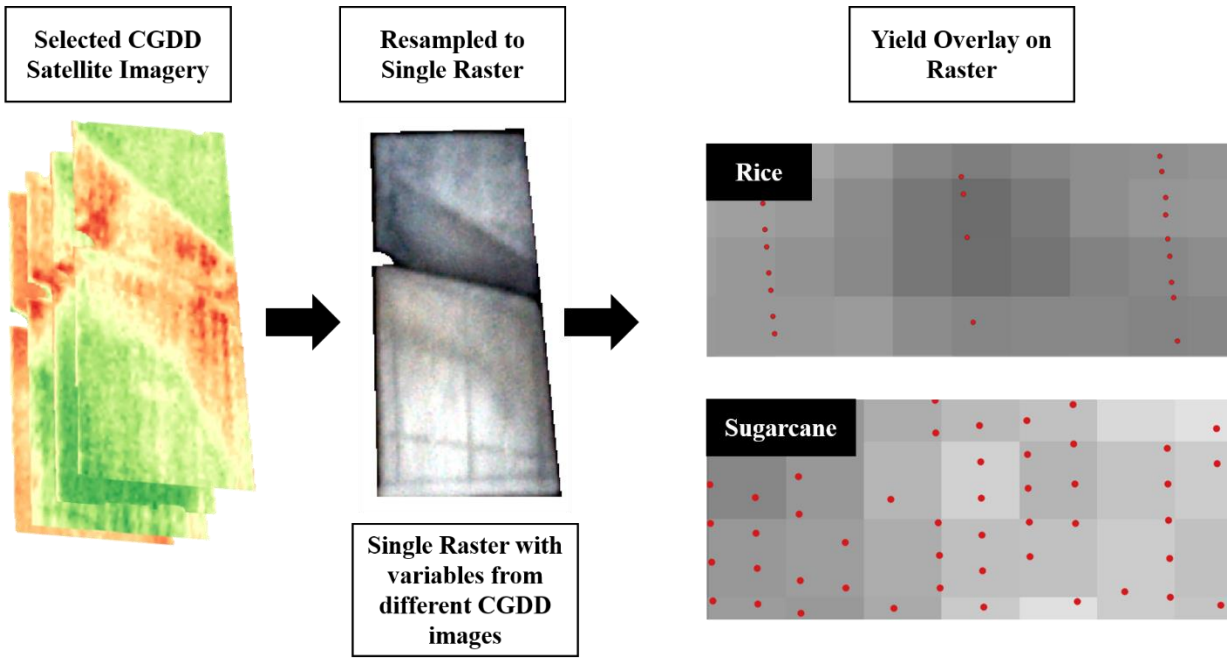

Figure 2.5. Example of resampling the multiple rasters from different Cumulative Growth Degree Days (CGDD) to a single raster and harvester yield overlaying process.

The binding process of yield and raster information was also performed using a customized script under the language R. The single raster image pixels were converted to polygons using the function *rasterToPolygons* from the package *raster*, then using the function *st_intersection* from the package *sf* the yield points were attached to an identification number to each pixel. In sequence, the points with the same identification numbers were averaged and the obtained values added to its corresponding pixel. Consequently, in the end of this process a single file with the satellite single bands and calculated VIs for all the selected CGDD plus harvester yield was obtained.

- Yield Map Prediction Model Generation

The framework proposed focused on the usage of ML algorithms to create the prediction model. Multiple machine learning algorithms are available and could be applied to the data to predict yield. However, it is important to notice that a few studies had indicated that Random Forest Regressors (RFR) tend to present a better performance for yield prediction using remote sensing (Fu et al., 2020; Ramos et al., 2020; Sanches et al., 2018), algorithm adopted for the development of this study.

Random Forests (RF) are ensemble machine learning algorithm that can be used for regression and classification. A RF is basically an ensemble of Decision Trees, which is another machine learning algorithm. Each tree within a RF is built by using a statistical method called bagging, in which samples are taken from the dataset to build each tree (Figure 2.6). The bagging process is developed randomly and repeatedly, so each observation has a similar probability of being selected. If bootstrap technique is used, this sampling process is performed with replacement, which means that the observation can be present in more than one sample. In addition, for the determination of a node split an extra-randomness is added by performing the best feature search in a random subset of the features (the number of features used in each subset is a parameter defined by user through a maximum number of features selection model).

After building each tree the decisions recorded and through an ensemble technique – averaging for regression and voting/weighting for classification – the final result is obtained (Breiman, 2001; Géron, 2017; Hastie, Tibshirani, & Friedman, 2009). Usually, within the RF the number of trees and their depth (number of nodes) are two main parameters that can be set by the user to obtain the best prediction model. These two parameters are responsible for the determination of the complexity of the final model. The tree depth parameter needs special

attention, since depending on the dataset used, increasing the tree depth can result in the model overfit.

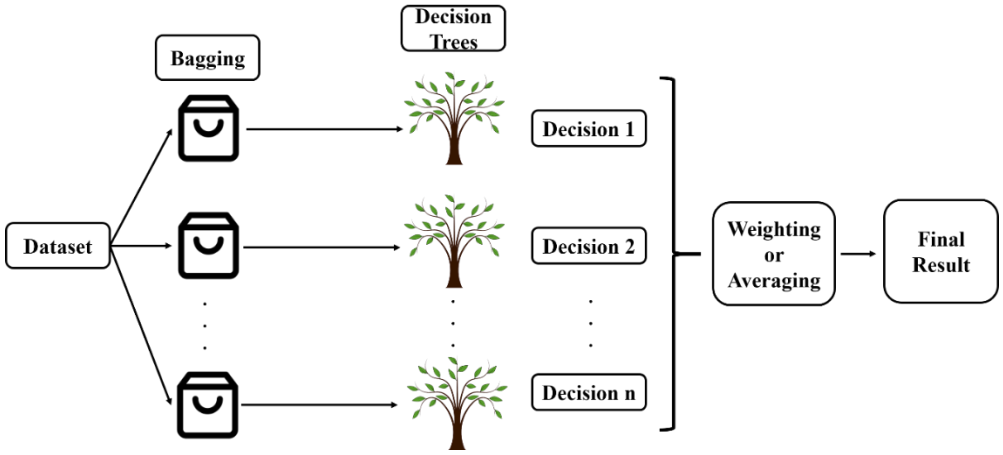

Figure 2.6. Random forest structure and running procedure to perform predictions – n is a maximum number of trees set by the user

To evaluate the proposed framework, a random forest regression model was trained and tested using 12 variables for each CGDD selected (imagery single bands red, green, blue and near infra-red plus the calculation of vegetation indices presented on Table 2.1). The RFR was built using the package *scikit-learn* (Pedregosa et al., 2012) on a customized script under the language Python 3. According to *scikit-learn* description of the function *RandomForestRegressor* – used in this study for the model development - a few parameters can be set by the user to obtain the best performance model. From those parameters the following were adjusted in the development of this study: number of trees (also known as number of estimators), maximum depth of a tree, minimum number of samples to split an internal node, minimum number of samples at the leaf node, method to determine the maximum number of features for the best split and if bootstrap is performed or not while building the trees (if bootstrap is not used, the whole dataset is used to build the trees).

To adjust the best parameters for the Radom Forest, a process called hyperparameter tuning was performed. In this study the package *Optuna* (Akiba et al., 2019) was used in order to efficiently obtain the best parameters for the yield prediction models generation for rice and sugarcane. *Optuna* is a software framework scripted under Python language with the objective of automatic machine learning hyperparameter optimization. It used a Bayesian optimization algorithm that allow an efficient selection of the best performing parameters by recording the performance of previous sets of parameters and usage of an early trial pruning system (Akiba et al., 2019).

To perform the hyperparameter tuning using *Optuna*, the number of trials (interactions), a search space and an objective function needs to be set. The search space for this study was set according to the parameters presented on Table 2.3. The objective function used was a RFR which had its performance evaluated according to the coefficient of determination R^2 of the prediction (also known as score), obtained from the RFR *scikit-learn* function (Equation 1). The number of trials used were equal to 100.

$$R^2 = 1 - \frac{\sum_{i=0}^{n_{obs}-1}[(y_i - \hat{y}_i)^2]}{\sum_{i=0}^{n_{obs}-1}\left[\left(y_i - \frac{1}{n_{obs}}\sum_{i=0}^{n_{obs}-1} y_i\right)^2\right]} \quad (Eq.1)$$

Where

$\hat{y}_i$ is the yield predicted by the model

y_i is the yield assessed by the harvester

n_{obs} is the total number of observations

Table 2.3. Search-space used as input for *Optuna* hyperparameters optimization – parameters were selected according to *scikit-learn RandomForestRegressor* parameters description

Parameter	**Data Type**	**Search Space**
Number of Estimators	Integer	10 to 1500
Maximum Depth	Integer	2 to 20
Minimum Samples Leaf	Integer	1 to 5
Minimum Samples Split	Integer	2 to 6
Maximum Feature	Categorical	auto, sqrt, log2
Bootstrap	Boolean	True or False

To generate the model, RFR requires a set of data for the training. In addition to evaluate the results it is important to also have a testing dataset. Therefore, the dataset obtained in the final assemble was split into train and test sets. According to Ferraciolli, Bocca, & Rodrigues (2019) one of the biggest mistakes when using regression machine learning models for spatial data is related to the train and test dataset split. Usually, the split is done by randomly separating the data as 80% for training and 20% for test. However, regression models do not account for spatial autocorrelation. Therefore, when performing the evaluation of the results, the errors might be sub estimated. This situation would happen if some of the data selected for testing had spatial autocorrelation to data selected for training, creating a bias. Thus, to reduce the spatial autocorrelation, the dataset split occurred by randomly selecting two fields for training and one field for testing for each crop. Fields RC1 and RC2 (19,841 observations), SC1 and SC2 (14,812 observations) were used for training for rice and sugarcane, respectively. While RC3 (4673 observations) and SC3 (8408 observations) were used for testing (Figures 2.1 and 2.2).

- Results Analysis

To analyze the performance of the ML model developed based on the framework proposed and to perform the parameter optimization, a k-fold cross validation (CV) was performed. A k-fold Cross Validation is a resampling method in which a set of data is split in k number of folds or groups. Then the first fold is considered as a validation set while the

remaining of the data is used to fit the studied model. Then, the errors of the predictions are calculated based on the hold-out fold kept for validation. This process is repeated k times, always using different observations as the validation set (James et al., 2013). In this study, a 10-fold CV was performed only when using the training dataset.

During the hyperparameter optimization, the 10-fold CV was performed for each of the interactions, and the coefficient of determination was evaluated at each fold. The average of the coefficient of determination for each fold took place, was reported, and used as the value to be optimized while changing the RFR hyperparameters. In addition, using the best hypermeters obtained from *Optuna*, the 10-fold CV was performed to also evaluate the final model. Therefore, at each fold the average error (AE - Equation 2), mean average percentage error (MAPE – Equation 3), and the root mean squared error (RMSE – Equation 4) were obtained.

$$AE = \frac{1}{n_{obs}} \sum_{i=0}^{n_{obs}-1} |y_i - \hat{y}_i| \; (Eq. 2)$$

$$MAPE = \left[\frac{1}{n_{obs}} \sum_{i=0}^{n_{obs}-1} \left(\frac{|y_i - \hat{y}_i|}{y_i}\right)\right] * 100 \; (Eq. 3)$$

$$RMSE = \sqrt{\frac{1}{n_{obs}} \sum_{i=0}^{n_{obs}-1} (y_i - \hat{y}_i)^2} \; (Eq. 4)$$

Where

$\hat{y}_i$ is the yield predicted by the model

y_i is the yield assessed by the harvester

n_{obs} is the total number of observations

It is important to notice that, if during the CV, the splits are performed randomly, the same issue with spatial autocorrelation previously described, will occur. Therefore, one option to overcome this issue is to split the data in different groups, avoiding spatial autocorrelation among them, then perform a Group k-fold CV. Using this process, the same group is not represented on both training and validation sets in a same fold. Therefore, a built-in function called *GroupKFold* under the Python package *scikit-learn* was used. Figure 2.7 exemplifies how the *GroupKFold* works when a 4-fold CV is used. In this study, the same approach was used but using a 10-fold CV.

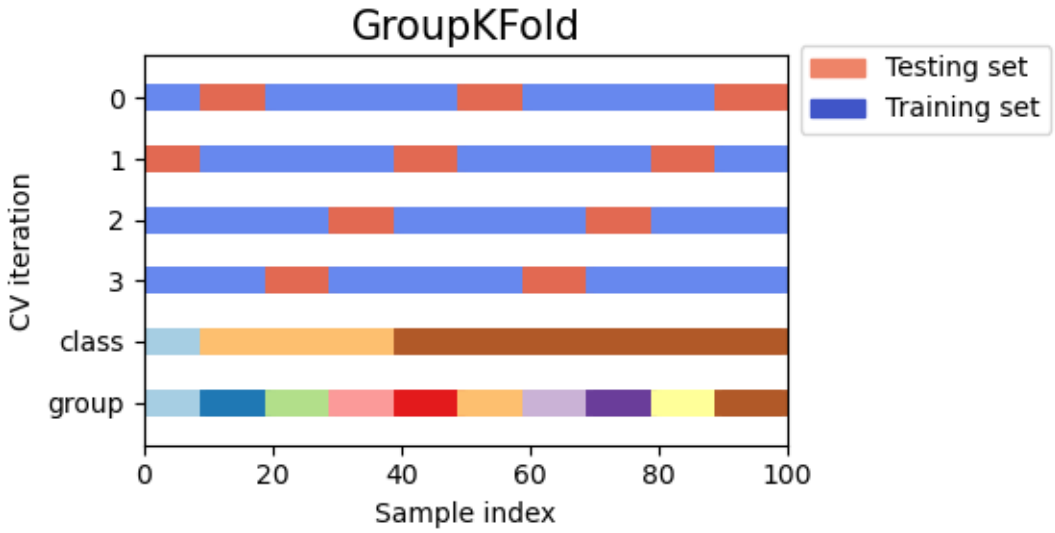

Figure 2.7. Example of a 4-fold group Cross Validation (CV) that is presented on *scikit-learn* manual (Pedregosa et al., 2012),; whilst a 10-fold group CV was used for the analysis of the yield map predictions.

In this scenario, ideally, there would be ten or more fields within the training dataset, which would be used as the groups for the folds. However, in the dataset used in this project only two fields were available within the training set. Since the *GroupKFold* function requires that the number of groups and folds should be at least the same, the training data had to be divided into

10 different groups to perform the 10-fold CV. Consequently, this split must be performed in a way to reduce the spatial autocorrelation. Thus, the methodology of a Spatial CV proposed by Ruß & Brenning (2010) was used. Using the procedure proposed by the authors, a *k*-means cluster algorithm was performed using a customized script under the Python 3 language in which the *KMeans* function from *scikit-learn* with a set number of 10 centers was applied in the Latitude and Longitude of the training sets for both crops. Figure 2.5 uses the rice training dataset to demonstrate the grouping clusters obtained from the *k*-means and further used as inputs in the *GroupKFold* function.

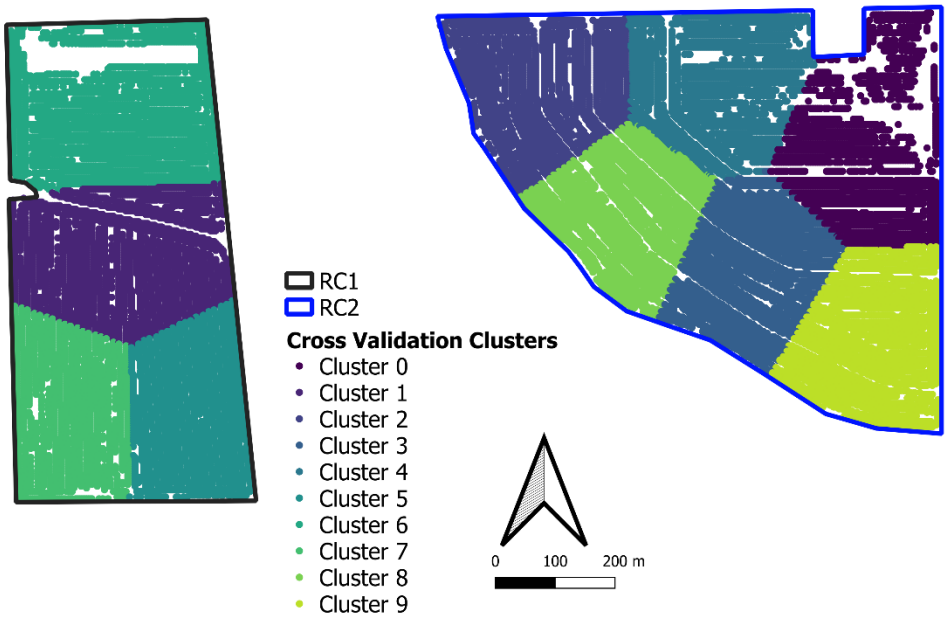

Figure 2.8. Rice training dataset clusters obtained from a *K*-means classification based on Latitude and Longitude. Clusters were used on the 10-fold Group Cross Validation

In addition, to evaluate the prediction model two maps for each testing field were generated to compare the spatial predictions obtained. One map represents the predicted values and the other the interpolated values for the filtered yield assessed by the harvester. In addition to the maps, descriptive statistics (DS) (mean, median, minimum, maximum, standard deviation, and coefficient of variation) for the interpolated and predicted yield of the fields were calculated.

DS and maps were obtained from a customized script built under the language R. Maps were generated using the R package *tmap* (Tennekes, 2018).

2.3. Results

2.3.1. Yield Filtering Process

Table 2.4 presents the descriptive statistics for the local filtering and raw yield data for rice and sugarcane fields. Based on the analysis of the mean and median obtained, after the filtering process, both values tend to be closer. Also, through the analysis of the maximum and minimum values before and after the filtering process, there is a reduction in the yield range, a characteristic related to the removal of outliers. In addition, a reduction in the standard deviation and coefficient of variation was observed on the filtered data compared to the unfiltered.

Table 2.4. Descriptive statistics for raw and filtered yield for all the study sites

Field	Mean	Median	Minimum	Maximum	SD*	CV**
SC1 - Raw	52.67	52.16	0.00	165.60	18.41	34.96
SC1 - Filtered	51.72	51.56	41.51	63.53	4.18	8.09
SC2 - Raw	56.03	53.65	0.00	328.43	23.27	41.52
SC2 - Filtered	51.63	51.52	42.52	61.53	3.53	6.85
SC3 - Raw	57.98	54.70	0.00	247.27	25.04	43.18
SC3 - Filtered	52.70	52.64	41.46	64.31	3.90	7.40
RC1 - Raw	7.7	7.6	0.9	13.5	2.0	25.9
RC1 - Filtered	7.0	7.0	3.7	9.8	0.9	12.9
RC2 - Raw	7.3	7.8	0.7	13.5	2.6	35.4
RC2 - Filtered	7.8	8.0	1.6	11.8	1.6	20.9
RC3 - Raw	8.1	8.3	0.7	13.5	2.1	25.6
RC3-Filtered	8.2	8.3	3.5	11.2	1.3	15.9

*SD – Standard deviation ** Coefficient of Variation – values presented in percentage. All other values are $Mg.ha^{-1}$

Figure 2.9 presents the maps for two fields, one from rice and the other from sugarcane. The figure presents maps for unfiltered (raw) and filtered data. Based on the raw data it is possible to observe and spot some high and low yield zones in the fields. However, multiple

erroneous data can also be observed, especially in the headlands of the fields or machine maneuvering spots. During the filtering process erroneous and outlier values were removed; therefore, in the filtered maps, the high and low yield zones become more evident.

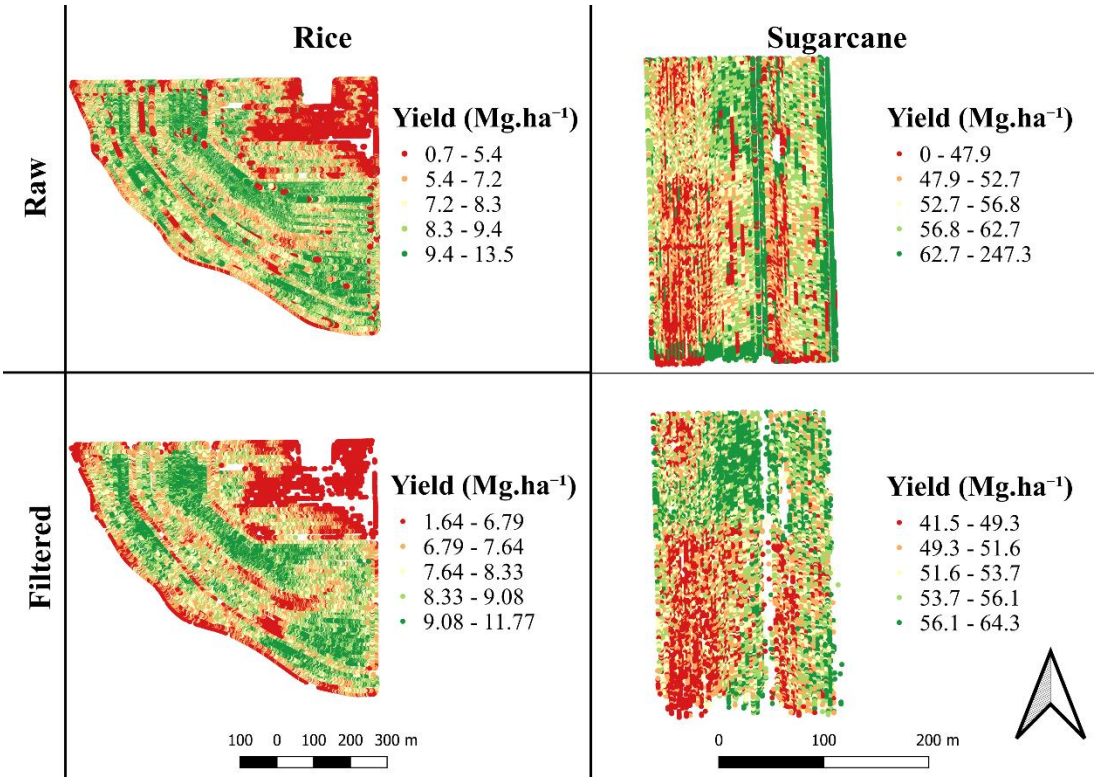

Figure 2.9. Comparison between the yield before and after the filtering process for fields RC2 and SC3

2.3.2. Imagery Organization and Final Dataset

Figures 2.10 and 2.11 present the number of images downloaded for each month within the production season of rice and sugarcane. The imagery downloads started on the month in which the first select CGDD was reached. Therefore, besides the fact that the sugarcane season started after the harvest that happened still in 2017, since the crop went through the winter season, it took until mid-March of 2018 to accumulate 115°C. Moreover, even though PS2 has a daily revisit frequency, there are some months in which not more than 2 images were downloaded. However, for both crops at least one image was available for the selected CGDDs.

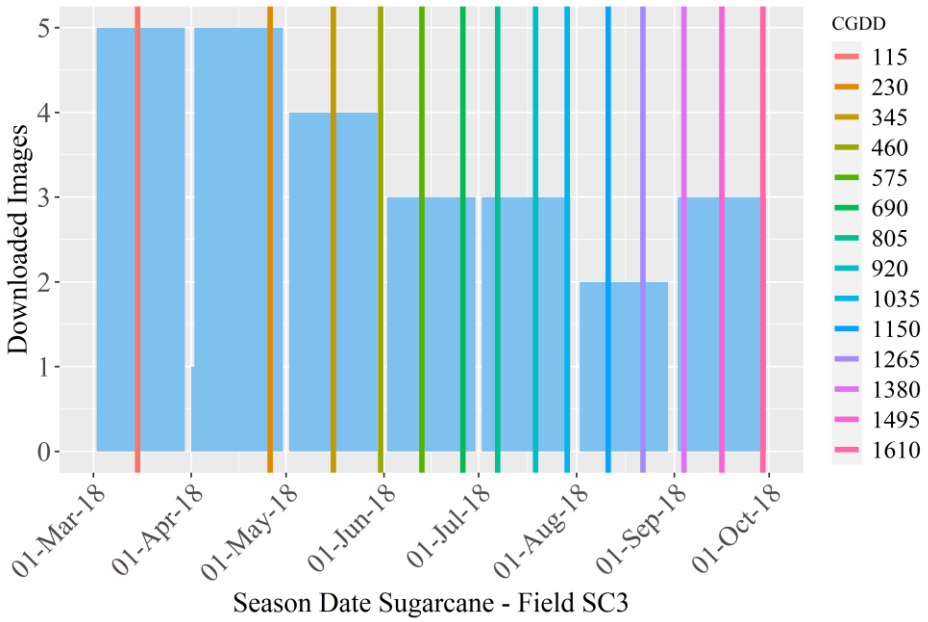

Figure 2.10. Number of downloaded satellite imagery for every month and selected CGDDs (°C day) through the sugarcane season Field SC3

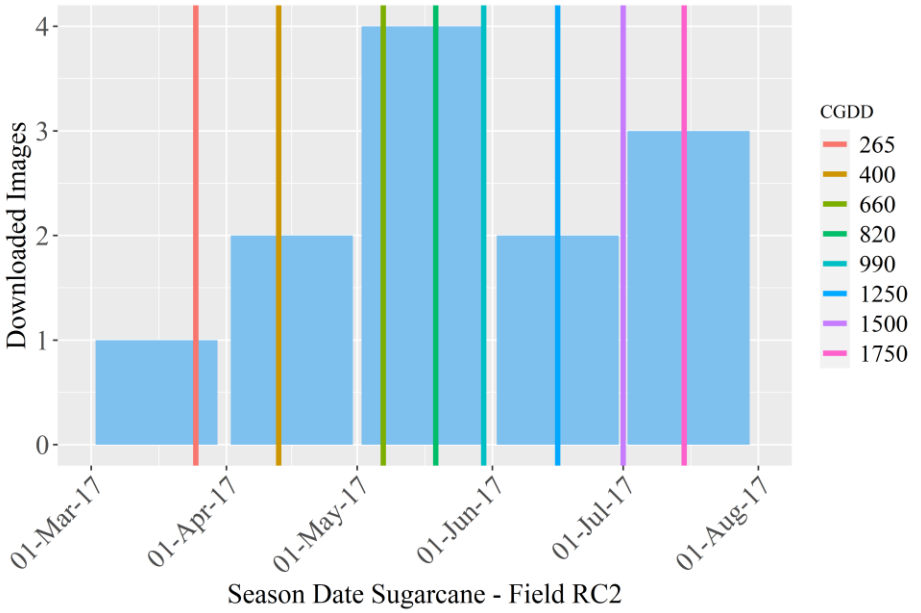

Figure 2.11. Number of downloaded satellite imagery for every month and selected CGDDs (°C day) through the rice season Field RC2

As described in the methodology, the filtered yield and the CGDD based organized images were combined. Resulting in final datasets with 24514 and 23220 observations with 97 and 169 variables for rice and sugarcane, respectively.

2.3.3. Yield Map Prediction Model Performance

Using the final dataset obtained by combination of the filtered yield and the satellite-based information, the prediction model was built. A different model was generated for each crop. The hyperparameters obtained from the optimization performed with *Optuna* are presented on Table 2.5.

Table 2.5. Best performing Random Forest Regression hyperparameters obtained from optimization using the package *Optuna*

Parameter	**Sugarcane**	**Rice**
Number of Estimators	681	461
Maximum Depth	9	8
Minimum Samples Leaf	1	1
Minimum Samples Split	2	2
Maximum Feature	"sqrt"	"sqrt"
Bootstrap	True	True

The results from the model prediction and the cross validation performed are presented in the following sections.

- Rice

Figure 2.12 presents the results obtained for each fold out of the 10-fold cross validation performed for the rice dataset during training. The error through the different folds are fairly constant varying from a maximum RMSE of 1.53 $Mg.ha^{-1}$ to a minimum of 0.57 $Mg.ha^{-1}$ with an average of 0.9 $Mg.ha^{-1}$. The average of the MAPE for the CV was approximately 11%.

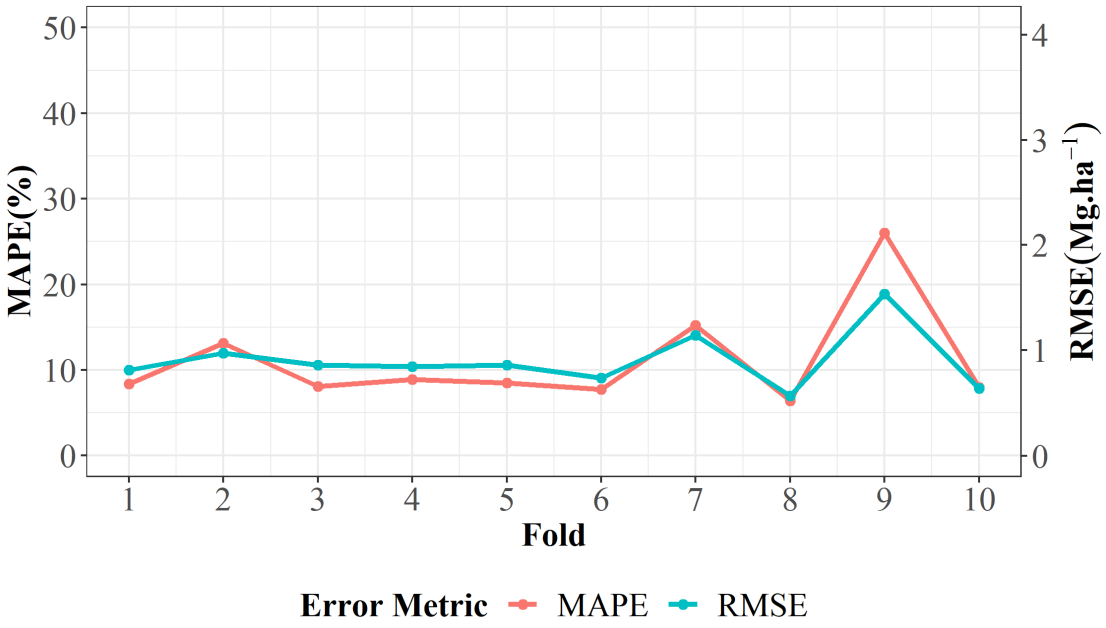

Figure 2.12. Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) for the 10-fold cross validation performed for the rice dataset during training.

After performing the CV, the final model took place based on all the data from RC1 and RC2. The obtainment of the final model was followed by the final test using the field RC3. The RMSE obtained for the test field was equal to 0.83 Mg.ha^{-1} and the MAPE was approximately 11%. Also, the predicted yield map for the RC3 field is presented on Figure 2.13. A visual analysis between the interpolated and predicted yield leads to the identification of similar high and low yield zones between the maps. In addition, the latest image used for the prediction was obtained around 10 days before the harvest date.

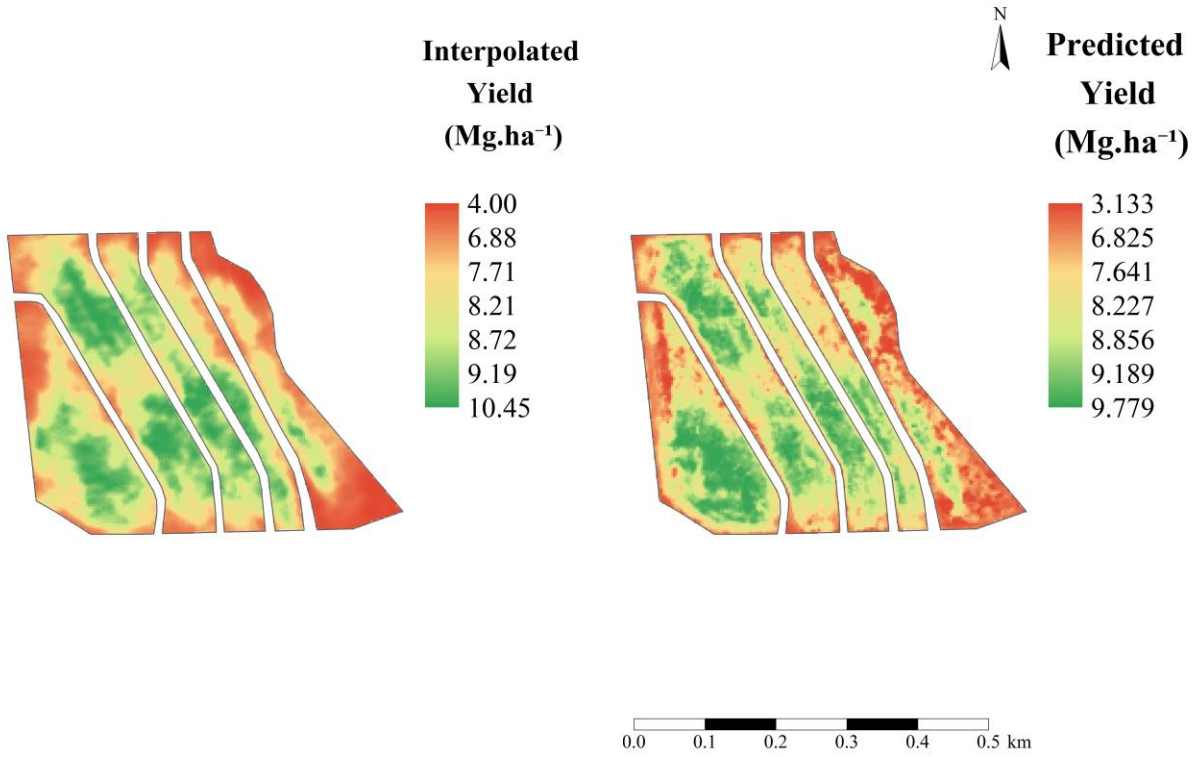

Figure 2.13. Interpolated (left) and Predicted (right) yield maps for rice in field RC3

Besides the error metrics and visual analysis, the descriptive statistics also presents a very close relationship between the interpolated and predicted yield data. As presented on Table 2.6, the average between the two maps differ in 0.01 $Mg.ha^{-1}$. In addition, the predicted yield maximum and minimum values were close to the interpolated yield.

Table 2.6. Descriptive statistics for the interpolated and predicted yield values for rice test site RC-3

Data	Mean	Median	Minimum	Maximum	SD*	CV**
Yield Interpolated	8.06	8.23	4.00	10.45	1.21	15.02
Yield Predicted	8.07	8.25	3.13	9.78	1.12	13.84

*SD – Standard deviation ** Coefficient of Variation – values presented in percentage. All other values are $Mg.ha^{-1}$

- Sugarcane

The results obtained for sugarcane have a similar trend to the results for rice. As Figure 2.14 presents, there is a slightly variance in the 10-fold cross validation error, with RMSE

varying from 2.73 Mg.ha^{-1} to 4.05 Mg.ha^{-1} and average of 3.14 Mg.ha^{-1}. The RMSE magnitude for sugarcane is higher than for rice because the yield magnitude of sugarcane is also higher (over 40 Mg.ha^{-1}) than for rice (8 Mg.ha^{-1}).

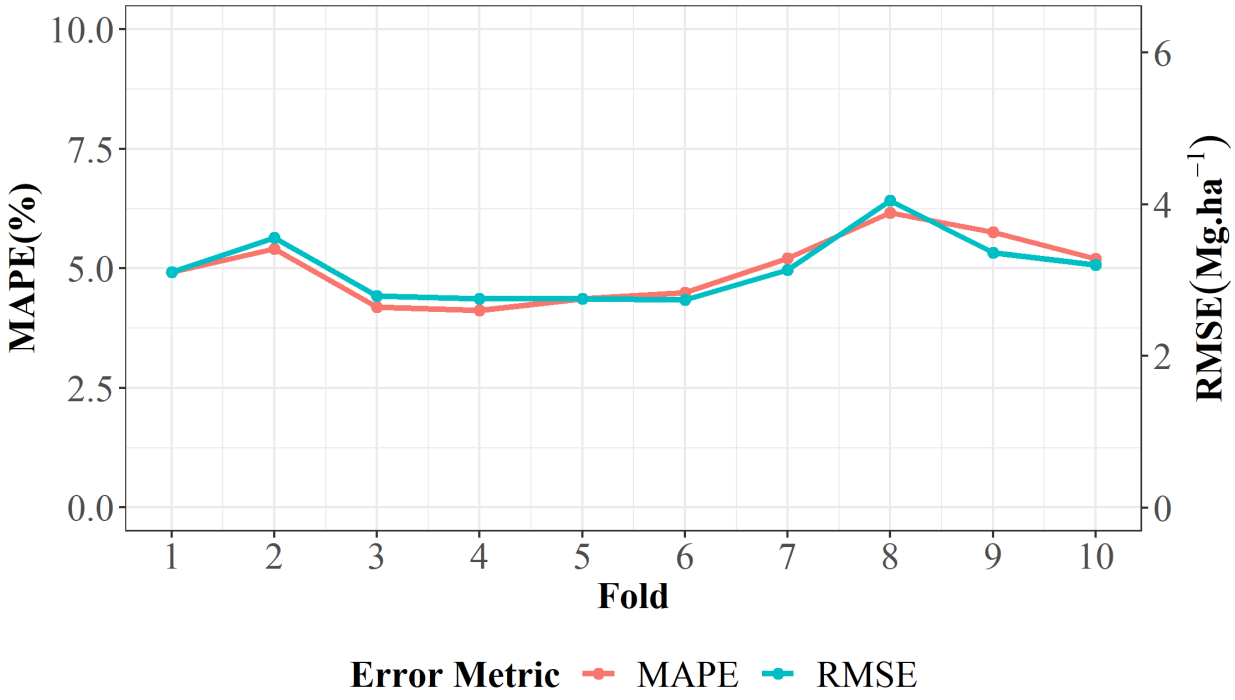

Figure 2.14. Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) for the 10-fold cross validation performed for the sugarcane dataset during training.

Such as for rice, after performing the CV, the training dataset (SC1 and SC2) was used to obtain the final model, which was applied to the test field, SC3. The test RMSE was 3.10 Mg.ha^{-1} and the MAPE approximately 4.8%. The accuracy of the prediction for this field is also represented in the predicted yield map presented in Figure 2.15. As for the rice maps, similar high and low yield zones can be identified for both maps, with few regions where a higher yield zone is identified in the interpolated map than in the predicted map. Also, the latest image used for the yield map prediction dataset of the test field was obtained 20 days before harvest.

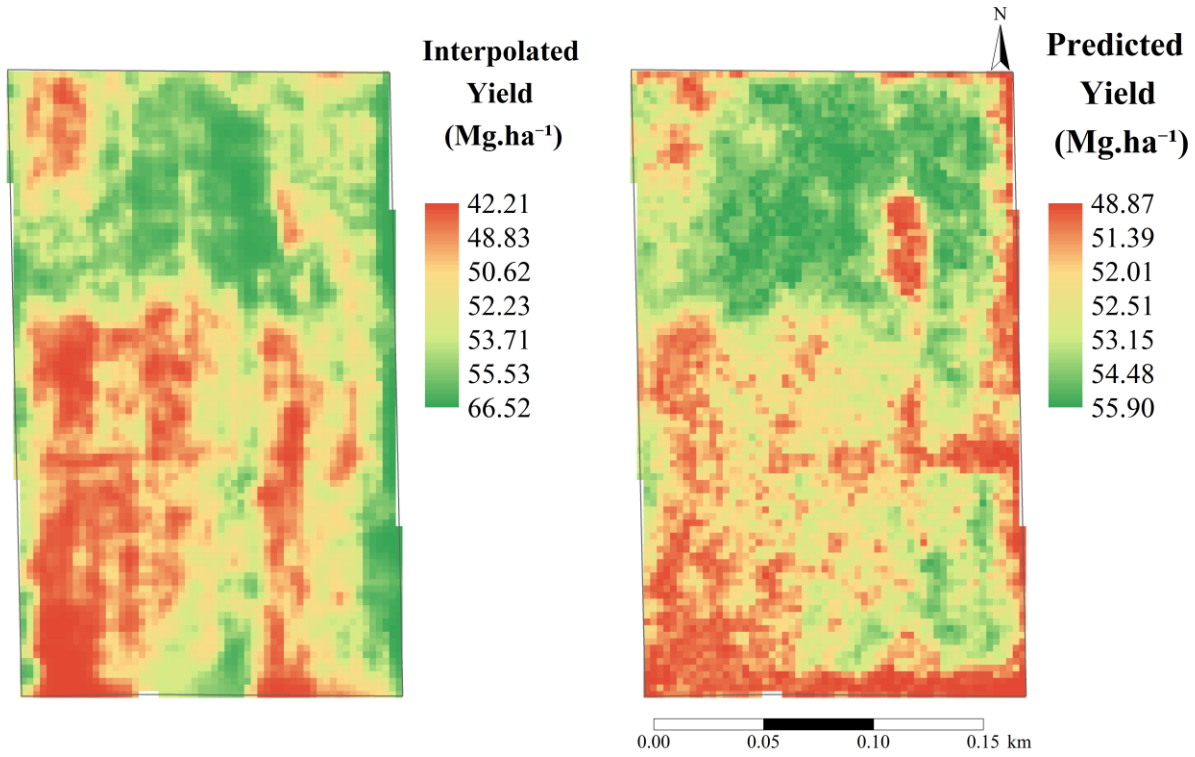

Figure 2.15. Interpolated (left) and Predicted (right) yield maps for sugarcane in field SC3

Table 2.7 presents the descriptive statistics for the interpolated and predicted yield maps. The difference in the average yield between the maps is equal to 0.42 Mg.ha^{-1}. However, the yield predicted has a lower variance than the mapped by the interpolated data. This condition can be observed through the lower coefficient of variation and the lower data range in the predicted yield.

Table 2.7. Descriptive statistics for the interpolated and predicted yield values for sugarcane test site SC-3

Data	Mean	Median	Minimum	Maximum	SD*	CV**
Yield Interpolated	52.34	52.27	42.21	66.52	3.54	6.76
Yield Predicted	52.76	52.53	48.87	55.90	1.39	2.63

*SD – Standard deviation ** Coefficient of Variation – values presented in percentage. All other values are Mg.ha^{-1}

2.4. Discussion

The yield filtering process adopted in this chapter generated similar results to previous studies that used similar methodologies to filter yield data. Maldaner et al. (2018), Spekken et al. (2013) and Sudduth, Drummond, & Myers (2012), after performing global and local filtering observed variations in average and reduction of the coefficient of variance of the data. Also, according to those authors, a removal of errors related to the harvesting process (maneuvers, speed fluctuation and others) was also observed, increasing the map accuracy. Therefore, the results obtained corroborate the need of yield filtering for the obtainment of accurate yield maps, and in consequence, to build a reliable yield map prediction model.

For crops which yield monitors are still not available or that has a low adoption the literature presents a few options for the yield assess of those crops. Colaço et al. (2020) proposed a methodology for yield assess for hand-picked crops. Schueller et al. (1999), proposed a low-cost yield mapping methodology for citrus. Wei et al. (2020) based on the carrot harvesting system was able to measure the yield and produce yield maps. Therefore, still for those crops, yield can be assessed and used as a target variable. Thus, the framework proposed in this study might not be limited to rice and sugarcane, or crops that have commercial yield monitors available.

Through the analysis of Figure 2.10 and 2.11 it is possible to observe that for a few months no more than 2 images were downloaded. This low amount of download images for a specific section of the crop season is due to the presence of a higher cloud cover then the set limit (10%). Considering that PS2 has a daily revisit frequency and still few images without cloud cover were available, for other satellites with a lower revisit frequency, the appearance of clouds on the images might be a limitation to the usage of this framework.

The yield prediction results obtained present high potential to the usage of satellite imagery in the obtainment of yield maps for rice and sugarcane. The RMSE values found during the CV are similar to values reported on previous research that evaluated the usage of remote sensing for rice yield prediction. Chang (2012) using a spectroradiometer obtained RMSE values of approximately 0.7 $Mg.ha^{-1}$ when using the spectral bandwidth relative to the blue, green, red and NIR bands to predict yield. Wan et al. (2020) using an Unmanned Aerial Vehicle (UAV) equipped with a multispectral camera, obtained RMSE values of around 0.71 $Mg.ha^{-1}$ when using a combination of VIs, canopy height and canopy coverage information obtained from the sensors embedded in the UAV. Kim et al. (2017) used high resolution satellite imagery (pixel size of 5m) in combination to rice crop model and obtained to a regional scale, an absolute percentage error of 8%, while in the test site used in this chapter a value of 11% was found, showing a good opportunity to use fewer parameters to predict yield compared to a high input crop simulation model approach.

The average sugarcane yield map prediction RMSE of 3.14 $Mg.ha^{-1}$ reported by the CV results, outperformed when compared to the RMSE described in the literature for field scale predictions obtained using satellite imagery. Bégué et al. (2010) using a maximum and integrated NDVI methods for SPOT timeseries obtained a RMSE of 13.2 and 15 $Mg.ha^{-1}$. Morel et al. (2014) coupling sugarcane crop models and remote sensing, obtained a minimum RMSE of 10.8 $Mg.ha^{-1}$. However, it is important to mention that both studies were developed using multiple sites and multiple years, as well as the sugarcane yield magnitude was higher than the ones presented by the dataset used in this study (Bégué et al., 2010 – average yield for study location varied from 60 to 90 $Mg.ha^{-1}$; Morel et al., 2014 – average yield for study sites was reported as 112 $Mg.ha^{-1}$). The low yield reported in the present study is common for Louisiana since after

the harvest (usually happens from October through December), the winter time does not allow a fast growth of the crop, which will start accumulating more GDD during spring season. The overall sugarcane season in Louisiana has a length of 8-9 months (Gravois et al., 2016). In addition, the fields used were all third ratoon crops.

Through the analysis of the interpolated and predicted yield maps obtained for the test sites for both crops, it is noticeable on Figure 2.8, that the predicted yield map for rice seems to present a higher resolution than the interpolated map. A possible explanation for this is that the rice combine used to harvest this test field had a width of approximately 9m, while the imagery pixel size is approximately 3.7m. Resulting in a higher resolution map when the prediction was performed based on satellite imagery. Yang et al. (2013) also observed the same pattern when predicting yield maps for sorghum. On the other hand, the same cannot be observed for sugarcane, which reaffirms that this difference in resolution might be related to the harvesting process. The sugarcane harvester configuration usually requires the process to be performed for each row, usually spaced by approximately 1.5m.

As mentioned in the introduction of this chapter, yield maps are substantial for the development of precision agriculture. Therefore, for the tested crops and based on the dataset used, there is a potential of using satellite imagery and the proposed framework to predict yield maps. Thus, for crops in which yield monitors are not available or its usage still is scarce (such as sugarcane) the type of methodology proposed and used in this chapter can support PA development. In addition, the predicted yield maps for both crops were generated before the actual harvest dates. Thus, this information has the potential to be used as a tool for harvest logistics as well as for starting the next crop management plans.

2.5. Conclusion

A framework for yield map prediction using satellite imagery was set and tested for two crops, sugarcane, and rice. The yield predicted maps for both crops were substantially accurate and presented similar errors to the ones reported by previous research that also evaluated the potential of remote sensing on the yield prediction.

Therefore, the proposed framework has the potential to predict yield maps. However, further research needs to be developed to evaluate the performance of this framework for larger datasets, multiple fields, and years.

Chapter 3. A Random Forest Approach to Evaluate Yield Map Prediction Using High-Resolution Remote Sensing Imagery Acquired at Different Growth Stages for Rice and Sugarcane

3.1. Introduction

In the previous chapter, a framework for using machine learning (ML) and high-resolution satellite imagery was presented. Rice and sugarcane were used for testing the procedures and obtainment of final prediction maps. The maps obtained were able to address high and low yielding zones within the fields, which presented the potential of the technique and framework proposed.

The prediction of yield maps can be an important tool for better and faster development of PA especially for crops, such as sugarcane, in which yield monitors are not available or its usage is still scarce (Corredo et al., 2020; Wei et al., 2020). However, as mentioned by Adamchuk et al. (2011), besides the high value of the information obtained from the yield map, its analysis provides a retrospective information and cannot provide any "spatial and temporal inconsistencies" within the production season. In this scenario, the prediction of yield maps during the season instead of only the end could bring insights and support in-season management strategies.

According to Schut et al. (2018), the prediction of yield maps can allow the within season management of fertilizers and even the evaluation of the crop response to the fertilize application, as well as the crop monitoring for other management operations such as pesticide application, pest control, irrigation, and others. Therefore, considering that PA goals involve increasing profitability, productivity, efficiency, and sustainability, being able to predict the crop response in yield during the growing season would help farmers in their decision-making activities.

For example, an important management activity that can impact rice and sugarcane yield, and that is also related to high cost and environmental effects, is nitrogen application (Amaral et al., 2018; Bégué et al., 2010; Harrell et al., 2011; Lofton & Tubaña, 2016). Therefore, since yield and nitrogen application are related, predicted yield maps could be used as a guide for the in-season nitrogen management, helping farmers to decide whether to apply or not more nitrogen and where to do it. In general, the predicted yield maps could guide and be used as an additional tool for the evaluation of the spatial and temporal variability within the crop season. In this scenario, this study aims to evaluate the potential of high-resolution satellite imagery and machine learning on the in-season yield map prediction focusing in the usage of those maps as a management tool for the development of PA strategies in rice and sugarcane.

3.2. Material and Methods

The same six fields, three of rice and three of sugarcane, presented on Chapter 2 were also used in this chapter (Figures 3.1 and 3.2). Therefore, for the studied sites description, refer to Chapter 2 section 2.2.1. In addition, the methodology adopted in this chapter is based on the framework proposed and tested on Chapter 2 sections 2.2.2 and 2.3.

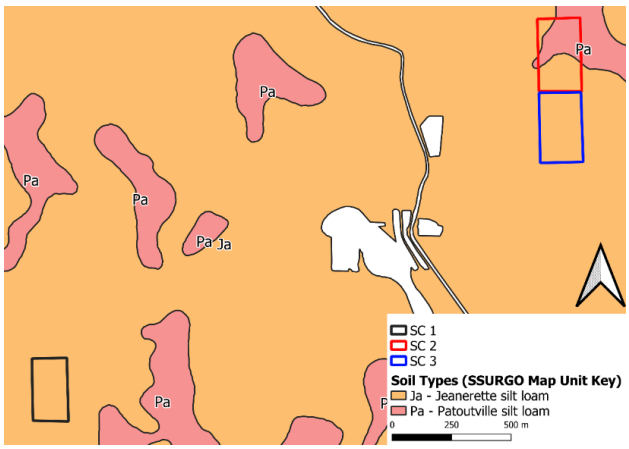

Figure 3.1. Soil characterization for the sugarcane fields located in Louisiana used for the evaluation of in season yield map prediction

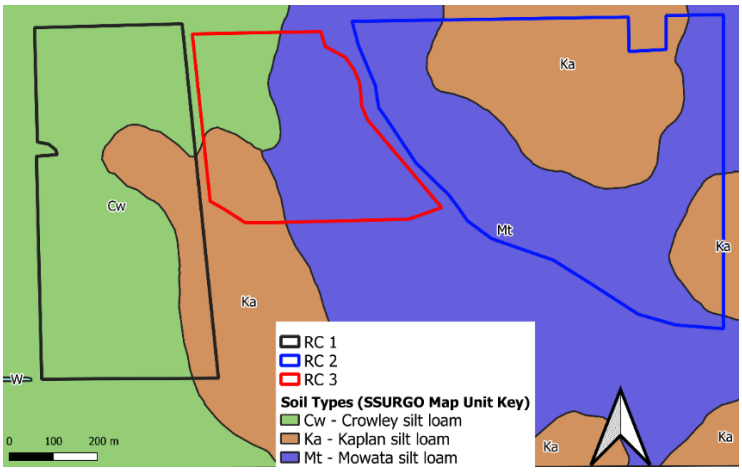

Figure 3.2. Soil characterization for the rice fields located in Louisiana used for the evaluation of in season yield map prediction

As a brief summary of the dataset, high-resolution satellite imagery from Planet Scope (pixel size of approximately 3.7m) were downloaded for each crop and for each field described on Figures 3.1 and 3.2. The images were then organized by selected cumulative growth degree days (CGDD) that were set according to the crop phenology stages. A total of 14 and 8 CGDDs were selected. All images were resampled to a single raster file and a series of vegetation indices (VIs – refer to Table 2.1) were calculated. In addition, yield data assessed by commercial yield monitors for rice and sugarcane were filtered and added to each crop respective final dataset.

Based on the final dataset, different scenarios were built for the evaluation of in-season yield map prediction. Those scenarios were developed to evaluate the accuracy of the predicted yield maps through the course of the crop growing season by accumulating satellite imagery. To evaluate the in-season prediction of yield maps, RFR was used. Thus, using the 14 and 8 different CGDDs for sugarcane and rice, respectively, from the beginning of the season towards the end, the CGDDs were added one by one to the scenarios. Table 3.1 and 3.2 presents the scenarios for rice and sugarcane, respectively.

Table 3.1. Rice different prediction scenarios used for the evaluation of the in-season yield map prediction. Each Cumulative Growth Degree Day represents a new image added to the dataset to perform the prediction

Scenario	Cumulative Growth Degree Days (°C day)
Rice 1	265
Rice 2	265 & 400
Rice 3	265, 400 & 660
Rice 4	265, 400, 660 & 820
Rice 5	265, 400, 660, 820 & 990
Rice 6	265, 400, 660, 820, 990 & 1250
Rice 7	265, 400, 660, 820, 990, 1250 & 1500
Rice 8	265, 400, 660, 820, 990, 1250, 1500 & 1750

Table 3.2. Sugarcane different prediction scenarios used for the evaluation of the in-season yield map prediction. Each Cumulative Growth Degree Day represents a new image added to the dataset to perform the prediction

Scenario	Cumulative Growth Degree Days (°C day)
Sugarcane 1	115
Sugarcane 2	115 & 230
Sugarcane 3	115, 230 & 345
Sugarcane 4	115, 230, 345 & 460
Sugarcane 5	115, 230, 345, 460 & 575
Sugarcane 6	115, 230, 345, 460, 575 & 690
Sugarcane 7	115, 230, 345, 460, 575, 690 & 805
Sugarcane 8	115, 230, 345, 460, 575, 690, 805 & 920
Sugarcane 9	115, 230, 345, 460, 575, 690, 805, 920 & 1035
Sugarcane 10	115, 230, 345, 460, 575, 690, 805, 920, 1035 & 1150
Sugarcane 11	115, 230, 345, 460, 575, 690, 805, 920, 1035, 1150 & 1265
Sugarcane 12	115, 230, 345, 460, 575, 690, 805, 920, 1035, 1150, 1265 & 1380
Sugarcane 13	115, 230, 345, 460, 575, 690, 805, 920, 1035, 1150, 1265, 1380 & 1495
Sugarcane 14	115, 230, 345, 460, 575, 690, 805, 920, 1035, 1150, 1265, 1380, 1495 & 1610

As presented on Figure 3.3, at each scenario, only its respective CGDDs were selected from the dataset, and a yield map prediction model was generated and evaluated. To build the predictions models, a customized script under the language Python 3 came in use.

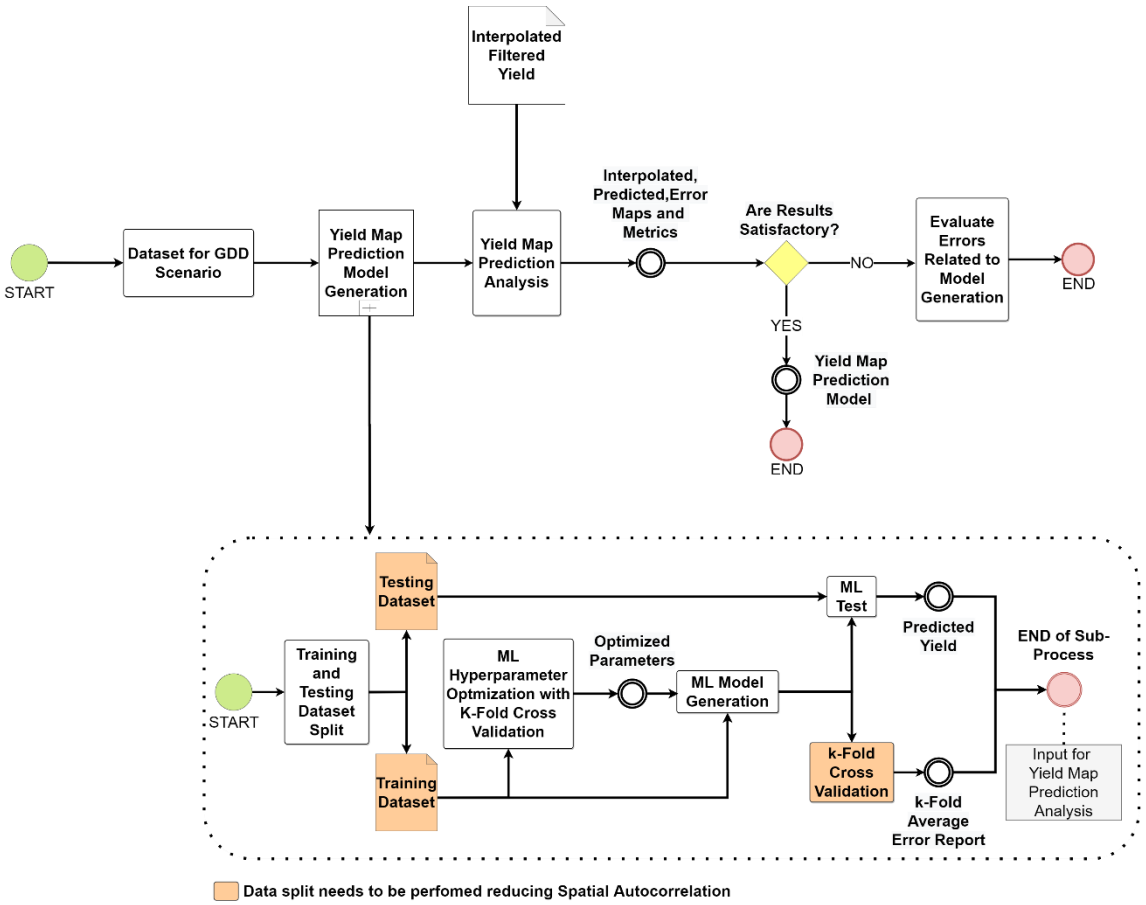

Figure 3.3. Yield map prediction model and evaluation flowchart

Then, the data was split into a training and testing dataset. To avoid spatial autocorrelation, the datasets for training and testing were not split randomly, but according to the fields available. Therefore, two fields for each crop were used for training and one was kept for testing. Fields RC1 and RC2, SC1 and SC3 were used for training for rice and sugarcane, respectively. While RC3 and SC2 were used for testing (Figures 3.1 and 3.2). Also, for the CV, the data from the training dataset was split following the methodology proposed by Ruß & Brenning (2010), also used in the previous chapter.

In sequence, for each scenario, the RF parameters were optimized using the Python package Optuna (Akiba et al., 2019), and the model was trained. To evaluate each scenario, root mean squared error (RMSE) and Pearson's correlation values between the predicted yield and

the one assessed by the harvester were recorded for each fold of a 10-fold Cross Validation (CV). In addition, using the optimized RF model on the test dataset, predictions for each scenario were generated and its results evaluated.

Based on the recorded results for the CV and test dataset, further analysis through a customized script built under the language R took place. A Scott-Knott test was carried out for the CV results, and for the test sites, maps were generated for each scenario and the descriptive statistics (mean, median, maximum, minimum, standard deviation and coefficient of variation) for the maps reported. The Scott-Knott (SK) analysis uses a hierarchical clustering algorithm to separate the treatment means into homogenous groups. Differently from other commonly used for treatment mean grouping, such as Tukey, through SK no overlaps among groups are obtained, which helps to distinguish the groups (Jelihovschi et al., 2014; Scott & Knott, 1974). The Scott-Knott test was performed within the R environment using the package *ScottKnott* (Jelihovschi et al., 2014).

In addition to the SK analysis, the filtered yield data for each test field was interpolated through ordinary kriging using the appropriate semi variograms. The interpolated map, was then used as a benchmark for comparison with the other maps, since this is the most common approach for yield map generation on PA (Blackmore & Moore, 1999; Isaaks & Srivastava, 1989).

3.3. Results

3.3.1. Sugarcane Yield Map Prediction Scenarios

Table 3.3 presents the average Person's correlation and RMSE for the 10-fold cross validation performed for each scenario and its corresponding SK group for the sugarcane dataset. According to the Scott-Knott results presented by this table, after the scenario Sugarcane 4, which considers the images related to CGDD from 115°C day to 460 °C day, there is no

significant difference among the scenarios at a probability of 5%. Besides that, the results still present an improvement on the model (increase of correlation and reduction of error) for further scenarios that include more images from the season. Sugarcane 12 scenario is the one with highest Pearson's correlation, and lowest prediction RMSE.

Table 3.3. Scott-Knott grouping of means for Person's correlation coefficient and RMSE for each sugarcane CGDD scenario 10-fold Cross Validation

Scenario	**Pearson's Correlation**	**RMSE ($Mg.ha^{-1}$)**
Sugarcane 1	0.179 b	3.354 a
Sugarcane 2	0.240 b	3.280 a
Sugarcane 3	0.184 b	3.448 a
Sugarcane 4	0.285 a	3.232 b
Sugarcane 5	0.325 a	3.194 b
Sugarcane 6	0.335 a	3.166 b
Sugarcane 7	0.304 a	3.188 b
Sugarcane 8	0.303 a	3.193 b
Sugarcane 9	0.313 a	3.182 b
Sugarcane 10	0.317 a	3.179 b
Sugarcane 11	0.336 a	3.146 b
Sugarcane 12	**0.342 a**	**3.128 b**
Sugarcane 13	0.335 a	3.142 b
Sugarcane 14	0.340 a	3.143 b

Means followed by different letters in the same column are significantly different by the Scott-Knott test at $\alpha = 0.05$. Bolded scenario presents the highest correlation and lowest RMSE

Figure 3.6 presents the interpolated yield map based on the harvester data (considered the best representation of yield in the field) and the predicted yield maps for the test field (SC2) at different scenarios. It is possible to notice that even though the Scott-Knott test presents a significant difference for the first three scenarios when compared the other eleven, all the maps for all scenarios present matching low and high yield regions with the interpolated map. However, through the season, the predicted maps tend to get more similar to the interpolated map. It is also important to notice that the range of the prediction usually is lower than the range

for the interpolated maps. This can be observed by the lowest appearance of dark red and green in the predicted maps and by the maximum and minimum values presented at Table 3.4.

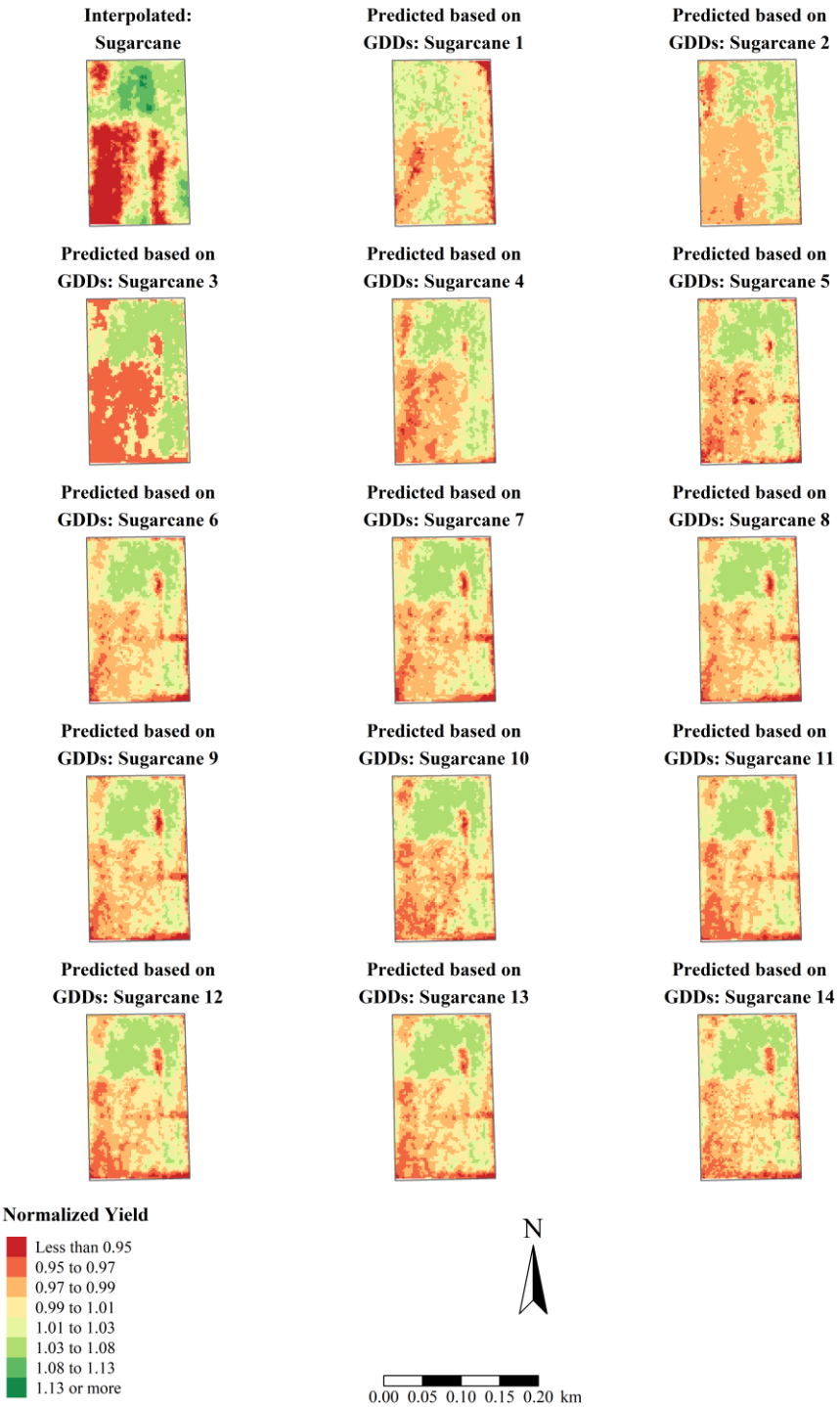

Figure 3.4. Sugarcane test field (SC2) interpolated and predicted normalized yield maps for the different CGDD scenarios

Table 3.4. Descriptive statistics for the sugarcane test site (SC2) maps at the different CGDD scenarios

Map	Mean ($Mg.ha^{-1}$)	Median ($Mg.ha^{-1}$)	Minimum ($Mg.ha^{-1}$)	Maximum ($Mg.ha^{-1}$)	SD* ($Mg.ha^{-1}$)	CV** (%)	Pearson's Correlation	RMSE ($Mg.ha^{-1}$)
Interpolated Yield	52.71	52.83	43.90	60.86	2.96	5.62	1.00	-
Sugarcane 1	52.70	52.78	46.36	55.12	1.25	2.37	0.30	2.85
Sugarcane 2	53.03	52.67	49.65	55.21	1.20	2.27	0.54	2.54
Sugarcane 3	53.22	52.85	50.65	55.44	1.70	3.19	0.64	2.34
Sugarcane 4	53.11	52.91	49.88	55.84	1.33	2.50	0.66	2.35
Sugarcane 5	52.87	52.57	48.70	56.31	1.38	2.61	0.62	2.37
Sugarcane 6	52.79	52.58	48.91	55.22	1.27	2.40	0.59	2.44
Sugarcane 7	52.83	52.61	48.84	55.95	1.39	2.63	0.61	2.39
Sugarcane 8	52.76	52.52	49.12	55.46	1.32	2.51	0.61	2.39
Sugarcane 9	52.79	52.56	48.35	55.62	1.37	2.60	0.60	2.40
Sugarcane 10	**52.55**	**52.28**	**48.90**	**55.89**	**1.46**	**2.79**	**0.66**	**2.29**
Sugarcane 11	52.70	52.48	49.39	55.50	1.39	2.63	0.65	2.31
Sugarcane 12	52.74	52.51	49.17	55.65	1.41	2.67	0.64	2.33
Sugarcane 13	52.77	52.55	49.16	55.49	1.35	2.57	0.64	2.35
Sugarcane 14	52.78	52.53	48.93	55.90	1.37	2.59	0.61	2.39

*SD – Standard deviation ** Coefficient of Variation – values presented in percentage. Pearson's correlation and RMSE were calculated based on the interpolated map. Bolded scenario presents the highest correlation and lowest RMSE.

Table 3.4 presents the descriptive statistics for the maps presented in Figure 3.6. In addition, the RMSE and Pearson's correlation were calculated for the predicted maps taking the Interpolated map as actual yield. Such as observed in the Figure 3.6, the range presented by the predicted maps is smaller than the range of the interpolated map. On the other hand, for all the scenarios, the mean of the predicted maps is similar to the mean of the interpolated map. According to Table 3.4 the lowest error and highest correlation came from the scenario with images from CGDDs 115, 230, 345, 460, 575, 690, 805, 920, 1035 and 1150°C (Sugarcane 10 - bolded in Table 3.4).

3.3.2. Rice Yield Map Prediction Scenarios

Table 3.5 presents the Pearson's correlation and RMSE averages computed during the Scott-Knott test performed for the 10-fold CV for each CGDD scenario for rice. According to the mean grouping presented by the Scott-Knott test, Rice 7 and 8 scenarios are statistically different from the remaining scenarios at a 95% confidence level. Therefore, this result indicates that the yield maps obtained at these two scenarios differ from the others, also that using imagery from the whole season might be necessary to obtain better accuracy on the prediction of yield maps for rice.

Table 3.5. Scott-Knott grouping of means for Person's correlation coefficient and RMSE for each rice CGDD scenario 10-fold Cross Validation

Scenario	**Pearson's Correlation**	**RMSE ($Mg.ha^{-1}$)**
Rice 1	0.22 d	1.20 a
Rice 2	0.25 d	1.22 a
Rice 3	0.20 d	1.23 a
Rice 4	0.32 c	1.15 a
Rice 5	0.38 b	1.09 a
Rice 6	0.46 b	1.01 b
Rice 7	0.51 a	0.96 b
Rice 8	**0.58 a**	**0.90 b**

Means followed by different letters in the same column are significantly different by the Scott-Knott test at $\alpha = 0.05$. Bolded scenario presents the highest correlation and lowest RMSE.

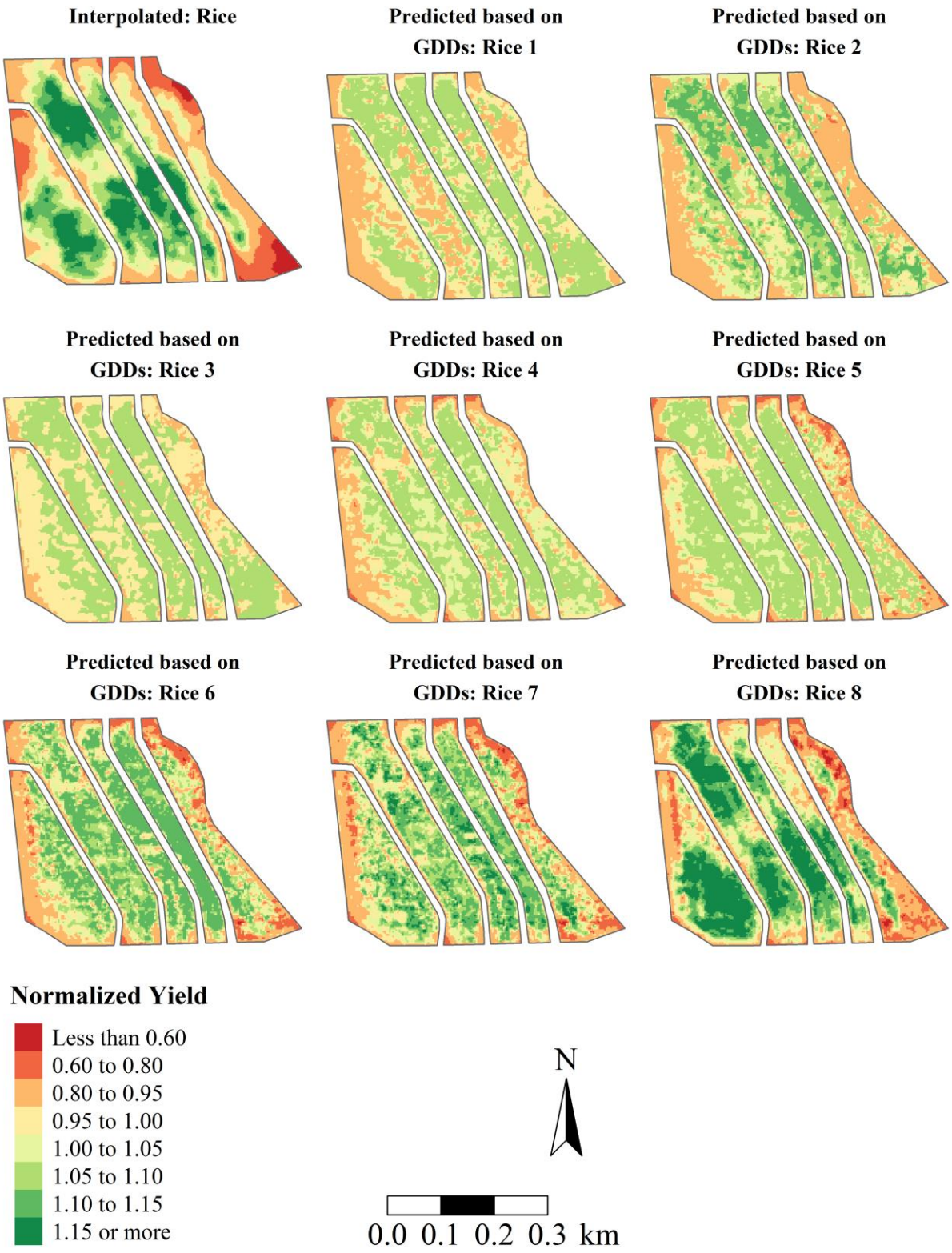

Figure 3.5 Rice test field (RC3) interpolated and predicted normalized yield maps for the different CGDD scenarios

Figure 3.7 presents the interpolated and predicted yield maps obtained for the rice test field (RC3). The maps comparison leads to a similar result as presented in Table 3.5 in which by increasing the number of images used from the season, the prediction accuracy also increases. Consequently, the predicted maps become similar to the interpolated towards the harvest date, until scenario Rice 8. Although the maps for the other scenarios do not represent the high and low zones as well as Rice 8, the low yielding zones are fairly constant throughout the season in all the maps.

Table 3.6 presents the descriptive statistics for the maps on Figure 3.7 and Pearson's correlation and RMSE for the predicted maps versus the interpolated (considered the actual yield). Comparing the averages of each map, the highest reported difference between the predicted and interpolated is for scenario Rice 1 at 0.33 $Mg.ha^{-1}$ and a RMSE of 1.27 $Mg.ha^{-1}$. On the other hand, at late ripening, when all images from the season were used for the yield map prediction (Rice 8 scenario), the RMSE is reduced to 0.71 $Mg.ha^{-1}$ with a mean difference of 0.01 $Mg.ha^{-1}$. Even though the mean difference decreases at later stages, there are several advantages to predict the final yield at early growth stages. For example, variable rate of fertilizers, fungicides, plant growth regulators, etc. In addition, the range of values predicted by the model are similar to the ones reported for the interpolated map.

Table 3.6. Descriptive statistics for the rice test site (RC3) maps at the different CGDD scenarios

Map	Mean ($Mg.ha^{-1}$)	Median ($Mg.ha^{-1}$)	Minimum ($Mg.ha^{-1}$)	Maximum ($Mg.ha^{-1}$)	SD* ($Mg.ha^{-1}$)	CV** (%)	Pearson's Correlation	RMSE ($Mg.ha^{-1}$)
Interpolated Yield	8.06	8.23	4.01	10.40	1.20	14.95	1.00	-
Rice 1	7.73	7.90	6.41	8.21	0.50	6.45	0.17	1.27
Rice 2	7.92	8.14	5.43	8.67	0.75	9.47	0.37	1.17
Rice 3	7.87	8.34	6.83	8.34	0.51	6.50	0.27	1.19
Rice 4	8.00	8.14	5.95	8.59	0.56	6.96	0.38	1.12
Rice 5	8.06	8.37	5.03	8.73	0.71	8.76	0.54	1.02
Rice 6	8.12	8.41	4.01	9.18	0.89	10.98	0.61	0.97
Rice 7	8.00	8.25	3.54	9.46	0.90	11.26	0.64	0.93
Rice 8	**8.07**	**8.25**	**3.13**	**9.78**	**1.12**	**13.84**	**0.82**	**0.71**

*SD – Standard deviation ** Coefficient of Variation – values presented in percentage. Pearson's correlation and RMSE were calculated based on the interpolated map. Bolded scenario presents the highest correlation and lowest RMSE.

3.4. Discussion

According to the results presented in this chapter and on Chapter 2, there is a potential for the usage of high-resolution satellite imagery and machine learning for sugarcane and rice yield map prediction. However, through the scenarios studied, sugarcane yield maps can address high and low yielding zones earlier in the season when compared to rice.

Therefore, from the point of view of in-season management, the sugarcane predicted yield maps for early CGDD scenarios have a potential application for guiding management strategies within the season. One possible usage of those maps is for splitting nitrogen application in sugarcane.

According to Gravois (2001), in situations when nitrogen split application in sugarcane is beneficial, one application should happen on early-April and the second at lay-by (usually before the crop reaches the tillering peak). Thus, the second application should happen no later than 800°C day (tillering peak happens when crop is in between 500°C – 800°C day; Inman-Bamber, 1994). Considering this information and the CV results presented on Table 3.3, a scenario that CGDDs are not higher than 800°C day could be used to help with the sugarcane in-season nitrogen application. To perform a simulation, Sugarcane 6 scenario model was chosen since it presented the best performance within the available models that match the requisites described in the previous sentence. In addition, the errors and Pearson's correlation on this scenario do not present statistical significance to the season's best performance model (Sugarcane 12) at 5% probability.

Thus, a *k*-means algorithm was used to classify the interpolated map, and the predicted maps for the scenarios Sugarcane 6 and 12 for the test field SC2 at two classes, higher and lower yield (Figure 3.8). In addition, a Cohen Kappa test was performed to assess the similarity of the maps. The results presented a moderate agreement at 95% confidence level when Sugarcane 6

and 12 scenarios were compared to the interpolated map. It is possible to observe that the clustered maps present similar matching zones. Thus, demonstrating the potential of using the map generated by the prediction model Sugarcane 6 (imagery from CGDDs 115, 230, 345, 460, 575 and 690°C day) as a guide for the split nitrogen application at lay-by. In addition, the predicted yield maps presented on Figure 3.6 could also be used to guide a variable rate application of ripeners.

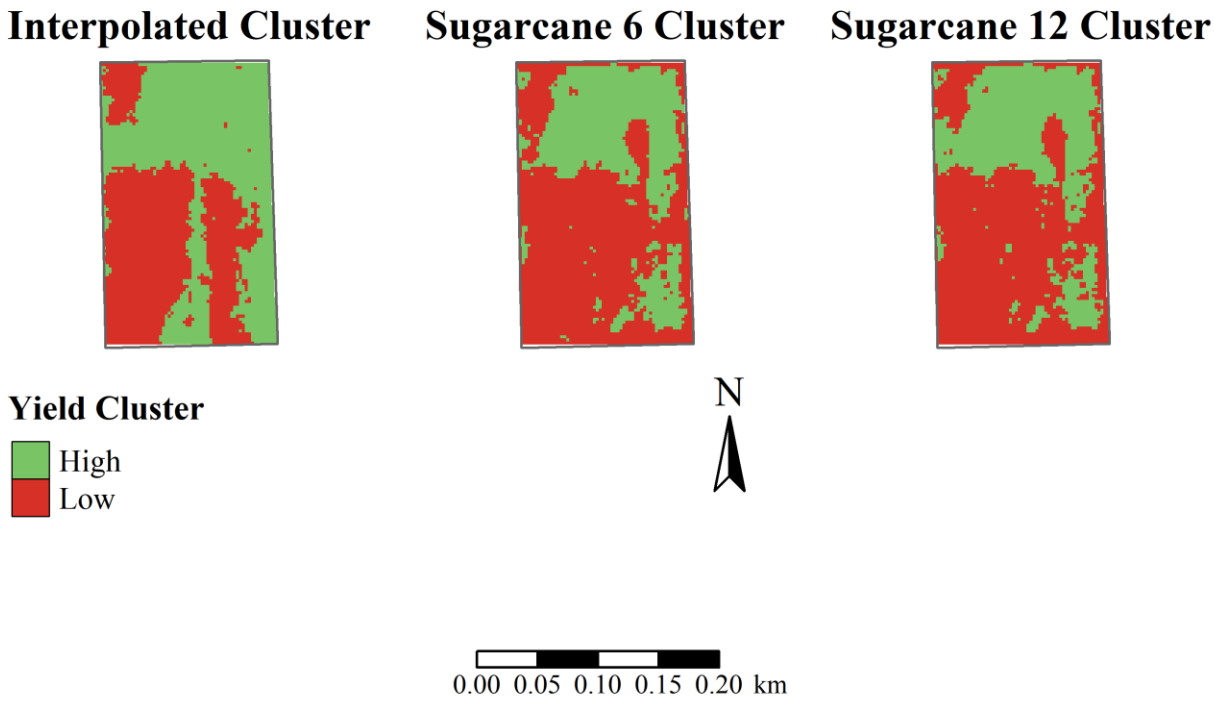

Figure 3.6. Two class *k*-means classification for interpolated and Sugarcane 6 and 12 scenarios predicted yield maps.

Besides that, it is important to mention that at scenario Sugarcane 6 the yield trend for the crop is not completely stablished. According to Bégué et al. (2010) this establishment would occur only at two months before harvest (around scenario Sugarcane 10). Therefore, that is the reason the same authors recommend that for sugarcane yield prediction based on vegetation index and on a single observation, satellite imagery from two months before harvest should be used. Two important observations: first for the present chapter, the scenarios used accumulated

the images over the development of the season instead of using a single image; second, besides no statistical significance is observed for the errors and correlations after scenario Sugarcane 4, the performance did increase until scenario Sugarcane 12, which aligns to the results and discussion reported by Bégué et al. (2010). In addition, the slightly increase in the error for the scenarios Sugarcane 13 and 14, could be related to the beginning of the sugarcane maturation or application of ripeners, an operation that is a standard in Louisiana sugarcane production ad often performed using herbicides (Orgeron et al., 2016) with the objective of increase the sucrose concentration. Therefore, this practice could affect how the later images can map the variability, causing the increase in the errors.

On the other hand, for rice, since the best performed yield map prediction was obtained only at the scenarios Rice 7 and 8 (early and late ripening) the usage of the maps for in-season management strategies becomes limited. Similar results have been reported by other researches which used remote and proximal sensors for yield prediction in rice. Guan et al. (2019) using an unmanned aerial vehicle (UAV) equipped with a multispectral camera, evaluated the potential of using a vegetation index for yield prediction in rice and concluded that imagery obtained early in the reproductive stage or in late ripening stage had the highest potential for yield prediction. Harrell et al. (2011) when evaluating a series of experimental trials in which a proximal sensor was also used for data collection, the best correlation between the sensor readings and yield was obtained when cultivars were beyond panicle differentiation. The same results were also obtained for other crops, such as corn. Shiratsuchi et al. (2011) using active canopy sensors in corn, found that just at late stages of crop development higher correlation values between VIs and yield are observed.

However, there are a few options for the usage of rice predicted yield maps. For example, since the maps are obtained before actually harvesting the fields, they could be used as a tool for planning the harvesting operations. Besides that, in southwest Louisiana, where the data used in this chapter was obtained, the climate condition and timing in which rice is produced allow the production of a second crop rice (ratoon or stubble rice). For this second crop, it is recommended that a nitrogen application should be performed five days after harvesting the first crop (Saichuk et al., 2014). In this scenario which the yield from the first crop can be obtained before the actual harvest – using the prediction model proposed in this study – this information could be used for decision making and planning the nitrogen application for the second harvest.

Therefore, the dataset used in this project presented a high importance for yield map prediction for both crops and beneficial results for the usage of the obtained information. On the other hand, more research needs to be developed using a larger number of fields and crop years to confirm that the results obtained in this present research can be generalized over the years, cultivars, varieties, crop age and others. Also, within machine learning techniques, deep-learning approaches, such as convolutions neural networks, could also be evaluated as possible model generation strategies.

3.5. Conclusion

Based on the results and dataset used, there is a potential for prediction of sugarcane yield maps based on high-resolution satellite imagery as early as 460°C day, when imagery since the beginning of the season at a step of 115°C day are used. Therefore, this information can be used for guiding and helping with in-season management. However, the maximum performance of the yield map prediction model used was obtained when imagery from the season until 1380°C day (about tillering stabilization) were used for prediction.

On the other hand, as presented by previous research, for rice yield map prediction, a better accuracy was obtained at late ripening, which limits the usage of this information for in-season management. Besides that, low error yield maps for rice were obtained before harvesting, which can be useful for nitrogen planning for second rice harvest in southwest Louisiana.

Chapter 4. Conclusions

Yield maps are an important tool for the development of PA. Therefore, the obtainment of this information is crucial to the development of management strategies and for decision-making. However, when yield monitors are available, farmers tend to not use the information to generate yield maps because of the need of in data pre-processing. In this scenario, the major goal of this project was to develop and evaluate a prediction system based on the usage of high-resolution satellite imagery and a machine learning algorithm.

To evaluate the proposed prediction system, two crops were used, rice and sugarcane. These crops were chosen because of their importance for the world food and energy security. In addition, those crops are important for the state of Louisiana and, in general, need more attention within PA. The main objective of this research was broken down to two specific objectives: the proposal of a framework for yield map prediction based on satellite imagery and the evaluation of different imagery dates or growth stage on those crops for yield map predictions.

The outcome from the first specific objective was a framework in which the final prediction system would only need the field boundary and season start date as inputs. Through the framework proposed, an interface could be built in which farmers would only have to add the boundary of the fields and the equivalent season start date for each field, then in the end of the season, a yield map would be generated. For the dataset used in this research the final maps obtained by the prediction system presented an accuracy similar to the ones presented by other researches in the literature that predicted yield for sugarcane and rice at field or regional scale. In addition, the generated maps presented similar low and high yield zones to the maps obtained from the yield data recorded by the harvesters.

For the second specific objective, the idea was to evaluate what was the potential of remote sensing and machine learning on the prediction of yield maps still during the season; therefore, farmers would be able to use this information to guide their management strategies. The dataset used in this research presented that for sugarcane, yield maps could be predicted early in the season. Thus, as an example, if farmers decided to do a split nitrogen application, yield maps could be obtained before the timeframe, in which a second nitrogen application is recommended to be done. Therefore, the predicted yield map could be precisely used to guide this second nitrogen application in a spatial manner. However, the in season yield map prediction could be used for many other management strategies based on the evaluation of anomalous regions within each field.

On the other hand, for rice, more accurate maps with similar high and low yield zones with the interpolated maps were just obtained at a later phenological stage. Therefore, this predicted information would have limited usages for in-season management strategies. However, yield maps would still be obtained before the actual crop harvest. Still a good information, since it could be used for planning harvesting. In addition, in Louisiana the weather condition and timing in which rice is planted allow a second harvest in rice crop. According to the state crop management recommendations, a nitrogen application should be performed five days after harvesting when a second crop is desired. In this matter, the predicted maps obtained from the first crop could be used to plan the second crop nitrogen application even before the first crop was harvested.

- Future Work

In general, the outcomes from this research presented a high potential for the usage of high-resolution satellite imagery and a machine learning algorithm for the yield map prediction

for rice and sugarcane. However, more studies need to be developed using a bigger database with more fields and multiple years in order to confirm the scalability of the results obtained in this study.

In addition, multiple variables could be added to improve the prediction models. Weather information, soil type, variety or hybrid are a few examples of variables that could help the system during the predictions, especially when considering multiple sites, years, and crops. Also, different machine learning algorithms, could be tested and evaluated. Since RF has a limitation on the extrapolation of the predictions out of the range of the data used for its training, algorithms that do not have this type of issue could be used. Some examples are deep learning techniques such as Convolution Neural Network and Recurrent Neural Network. One other option could be ensemble of different machine learning techniques; therefore, instead of using a single algorithm to obtain the yield map prediction, a group of different ML techniques could be used.

Therefore, it is possible to observe that the results presented in this study highlight the potential of the usage of the proposed framework techniques used. However, there is still a lot of question to be answered and a lot of opportunities of future research in this field of work.

References

Adamchuk, V. I., Rossel, R. A. V., Sudduth, K. A., & Schulze, P. (2011). Sensor Fusion for Precision Agriculture. In *Sensor Fusion - Foundation and Applications*. https://doi.org/10.5772/19983

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Ali, A., Martelli, R., Lupia, F., & Barbanti, L. (2019). Assessing multiple years' spatial variability of crop yields using satellite vegetation indices. *Remote Sensing*, *11*(20). https://doi.org/10.3390/rs11202384

Amaral, L. R., Molin, J. P., Portz, G., Finazzi, F. B., & Cortinove, L. (2014). Comparison of crop canopy reflectance sensors used to identify sugarcane biomass and nitrogen status. *Precision Agriculture*, *16*(1), 15–28. https://doi.org/10.1007/s11119-014-9377-2

Amaral, L. R., Molin, J. P., & Schepers, J. S. (2015). Algorithm for variable-rate nitrogen application in sugarcane based on active crop canopy sensor. *Agronomy Journal*, *107*(4), 1513–1523. https://doi.org/10.2134/agronj14.0494

Amaral, L. R., Trevisan, R. G., & Molin, J. P. (2018). Canopy sensor placement for variable-rate nitrogen application in sugarcane fields. *Precision Agriculture*, *19*(1), 147–160. https://doi.org/10.1007/s11119-017-9505-x

Azzari, G., Jain, M., & Lobell, D. B. (2017). Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment*, *202*, 129–141. https://doi.org/10.1016/j.rse.2017.04.014

Bégué, A., Lebourgeois, V., Bappel, E., Todoroff, P., Pellegrino, A., Baillarin, F., & Siegmund, B. (2010). Spatio-temporal variability of sugarcane fields and recommendations for yield forecast using NDVI. *International Journal of Remote Sensing*, *31*(20), 5391–5407. https://doi.org/10.1080/01431160903349057

Bengtsson, H. (2020). *future: Unified Parallel and Distributed Processing in R for Everyone*. Retrieved from https://github.com/HenrikBengtsson/future

Bhakta, I., Phadikar, S., & Majumder, K. (2019). *State-of-the-art technologies in precision agriculture : a systematic review*. (April). https://doi.org/10.1002/jsfa.9693

Birth, G. S., & McVey, G. R. (1968). Measuring the Color of Growing Turf with a Reflectance Spectrophotometer 1 . *Agronomy Journal*, *60*(6), 640–643. https://doi.org/10.2134/agronj1968.00021962006000060016x

Blackmore, S., Godwin, R. J., & Fountas, S. (2003). The analysis of spatial and temporal trends in yield map data over six years. *Biosystems Engineering*, *84*(4), 455–466.

https://doi.org/10.1016/S1537-5110(03)00038-2

Blackmore, S., & Moore, M. (1999). Remedial Correction of Yield Map Data. *Precision Agriculture*, *1*(1), 53–66. https://doi.org/10.1023/A:1009969601387

Bonnett, G. D. (2013). Developmental Stages (Phenology). In *Sugarcane: Physiology, Biochemistry, and Functional Biology* (pp. 35–53). https://doi.org/10.1002/9781118771280.ch3

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Cerri, D. G. P., & Magalhães, P. G. (2005). Sugar cane yield monitor. *2005 ASAE Annual International Meeting*, (October). https://doi.org/10.13031/2013.18878

Chang, W. (2012). Establishment of rice yield prediction model using canopy reflectance. *Journal of Computational Biology and Bioinformatics Research*, *4*(4), 39–50. https://doi.org/10.5897/jcbbr11.014

Chinosi, M., & Trombetta, A. (2012). BPMN: An introduction to the standard. *Computer Standards and Interfaces*, *34*(1), 124–134. https://doi.org/10.1016/j.csi.2011.06.002

Colaço, A. F., Trevisan, R. G., Karp, F. H. S., & Molin, J. P. (2020). Yield mapping methods for manually harvested crops. *Computers and Electronics in Agriculture*, *177*(August), 105693. https://doi.org/10.1016/j.compag.2020.105693

Corredo, L. D. P., Canata, T. F., Maldaner, L. F., & Lima, J. de J. A. de. (2020). Sugarcane Harvester for In-field Data Collection : State of the Art , Its Applicability and Future Perspectives. *Sugar Tech*. https://doi.org/10.1007/s12355-020-00874-3

da Silva, E. E., Rojo Baio, F. H., Ribeiro Teodoro, L. P., da Silva Junior, C. A., Borges, R. S., & Teodoro, P. E. (2020). UAV-multispectral and vegetation indices in soybean grain yield prediction based on in situ observation. *Remote Sensing Applications: Society and Environment*, *18*(March). https://doi.org/10.1016/j.rsase.2020.100318

Darr, M. J., Corbett, D. J., Herman, H., Vallespi-Gonzales, C., Dudas, B. E., & Badino, H. (2015). *Yield Measurement and Base Cutter Height Control Systems for a Harvester*.

Deliberto, M., Guidry, K., & Gravois, K. (2019). *Economic Importance of Louisiana Sugarcane Production in 2018*. Retrieved from https://www.lsuagcenter.com/~/media/system/2/6/2/7/26270a8030363b44684162366404 2158/02 summary section 2018 apdf.pdf

Dimitri, C., Effland, A., & Conklin, N. (2005). The 20th century transformation of U.S. agriculture and farm policy. In *Economic Information Bulletin* (Vol. 3). Retrieved from https://ageconsearch.umn.edu/record/59390/

Edmund, H., Chamberlain, S., & Ram, K. (2014). *rnoaa: NOAA climate data from R.* Retrieved

from https://github.com/ropensci/rnoaa

Eggleston, G., Legendre, B., & Tew, T. (2004). Indicators of freeze-damaged sugarcane varieties which can predict processing problems. *Food Chemistry*, *87*(1), 119–133. https://doi.org/10.1016/j.foodchem.2003.11.004

ESRI. (1998). *ESRI Shapefile Technical Description*.

FAO. (2009). *How to Feed the World in 2050*. Retrieved from http://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf

Fernandes, J. L., Rocha, J. V., & Lamparelli, R. A. C. (2011). Sugarcane yield estimates using time series analysis of spot vegetation images. *Scientia Agricola*, *68*(2), 139–146. https://doi.org/10.1590/s0103-90162011000200002

Ferraciolli, M. A., Bocca, F. F., & Rodrigues, L. H. A. (2019). Neglecting spatial autocorrelation causes underestimation of the error of sugarcane yield models. *Computers and Electronics in Agriculture*, *161*(December 2017), 233–240. https://doi.org/10.1016/j.compag.2018.09.003

Frankel, F. R. (1971). *India's Green Revolution: Economic Gains and Political Costs*. https://doi.org/10.2307/j.ctt13x0wb9

Fu, Z., Jiang, J., Gao, Y., Krienke, B., Wang, M., Zhong, K., … Liu, X. (2020). Wheat Growth Monitoring and Yield Estimation based on Multi-Rotor Unmanned Aerial Vehicle. *Remote Sensing*, *12*(3), 508. https://doi.org/10.3390/rs12030508

Fulton, J. P., Sobolik, C. J., Shearer, S. A., Higgins, S. F., & Burks, T. F. (2009). Grain Yield Monitor Flow Sensor Accuracy for Simulated Varying Field Slopes. *Applied Engineering in Agriculture*, *25*(1), 15–21.

Gao, L., Jin, Z., Huang, Y., & Zhang, L. (1992). Rice clock model-a computer model to simulate rice development. *Agricultural and Forest Meteorology*, *60*(1–2), 1–16. https://doi.org/10.1016/0168-1923(92)90071-B

Gatti, A., Naud, C., Castellani, C., & Carriero, F. (2018). Sentinel-2 Products Specification Document. *Thales Alenia Space*, 1–487.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly Media.

Gitelson, A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, *161*(2), 165–173. https://doi.org/10.1078/0176-1617-01176

Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment

in higher plant leaves. *Journal of Plant Physiology*, *160*(3), 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A. A., & Merzlyak, M. N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research*, *22*(5), 689–692. https://doi.org/10.1016/S0273-1177(97)01133-2

Gräler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-temporal interpolation using gstat. *R Journal*, *8*(1), 204–218. https://doi.org/10.32614/rj-2016-014

Gravois, K. (2001). *Sugarcane Production Handbook*.

Gravois, K. A., Zhou, M. M., Hoffmann, H. P., Piperidis, G., & Badaloo, G. (2016). Breeding new sugarcane varieties with enhanced ratooning ability. *Proceedings of the International Society of Sugar Cane Technologists*, *29*, 1683–1690.

Griffin, T. W., Yeager, E., & Ibendahl, G. (2019). Adoption of Precision Agriculture Technology. *22nd International Farm Management Congress*, *1*, 1–10. Tasmania, Australia.

Guan, S., Fukami, K., Matsunaka, H., Okami, M., & Tanaka, R. (2019). *Assessing Correlation of High-Resolution NDVI with Fertilizer Application Level and Yield of Rice and Wheat Crops using Small UAVs*. https://doi.org/10.3390/rs11020112

Haghverdi, A., Washington-Allen, R. A., & Leib, B. G. (2018). Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. *Computers and Electronics in Agriculture*, *152*(July), 186–197. https://doi.org/10.1016/j.compag.2018.07.021

Hamada, Y., Ssegane, H., & Negri, M. C. (2015). Mapping intra-field yield variation using high resolution satellite imagery to integrate bioenergy and environmental stewardship in an agricultural watershed. *Remote Sensing*, *7*(8), 9753–9768. https://doi.org/10.3390/rs70809753

Harrell, D. L., Tubaña, B. S., Walker, T. W., & Phillips, S. B. (2011). Estimating Rice Grain Yield Potential Using Normalized Difference Vegetation Index. *Agronomy Journal*, *103*(6), 1717–1723. https://doi.org/10.2134/agronj2011.0202

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. In *Springer Series in Statistics*. https://doi.org/10.1007/978-0-387-84858-7

Henebry, G., Viña, A., & Gitelson, A. (2004). The wide dynamic range vegetation index and its potential utility for gap analysis. *Papers in Natural Resources*, 50–56. Retrieved from http://digitalcommons.unl.edu/natrespapers/262/

Hengl, T. (2007). *A Practical Guide to Geostatistical Mapping of Environmental Variables*.

Hengl, T. (2009). *A Practical Guide to Geostatistical Mapping*. Retrieved from http://spatial-analyst.net/book/

Herder, G. Den, Van Isterdael, G., Beeckman, T., & De Smet, I. (2010). The roots of a new green revolution. *Trends in Plant Science*, *15*(11), 600–607. https://doi.org/10.1016/j.tplants.2010.08.009

Higgins, V., Bryant, M., Howell, A., & Battersby, J. (2017). Ordering adoption: Materiality, knowledge and farmer engagement with precision agriculture technologies. *Journal of Rural Studies*, *55*, 193–202. https://doi.org/10.1016/j.jrurstud.2017.08.011

Hijmans, R. J. (2020). *raster: Geographic Data Analysis and Modeling*. Retrieved from https://cran.r-project.org/package=raster

Holland, K. H., & Schepers, J. S. (2010). Derivation of a Variable Rate Nitrogen Application Model for In-Season Fertilization of Corn. *Agronomy Journal*, *102*(5), 1415–1424. https://doi.org/10.2134/agronj2010.0015

Huete, A., Didan, K., Miura, T., Rodriguez, E. ., Gao, X., & Ferreira, L. . (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Inman-Bamber, N. G. (1994). Temperature and seasonal effects on canopy development and light interception of sugarcane. *Field Crops Research*, *36*(1), 41–51. https://doi.org/10.1016/0378-4290(94)90051-5

International Society for Precision Agriculture (ISPA). (2019). ISPA Forms Official Definition of "Precision Agriculture." Retrieved October 8, 2020, from https://www.precisionag.com/market-watch/ispa-forms-official-definition-of-precision-agriculture/

IRRI, AfricaRice, & CIAT. (2010). *Global Rice Science Partnership (GRiSP)*. Retrieved from http://irri.org/our-science/global-rice-science-partnership-grisp

Isaaks, E. H., & Srivastava, R. M. (1989). *Applied Geostatistics*. NY: Oxford Univeristy Press, Inc.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Resampling Methods. In *An introduction to Statistical Learning*. https://doi.org/10.1007/978-1-4614-7138-7

Jelihovschi, E., Faria, J. C., & Allaman, I. B. (2014). ScottKnott: A Package for Performing the Scott-Knott Clustering Algorithm in R. *TEMA*, *15*(1), 003. https://doi.org/10.5540/tema.2014.015.01.0003

Jiang, Q., Fang, S., Peng, Y., Gong, Y., Zhu, R., Wu, X., … Liu, J. (2019). UAV-based biomass estimation for rice-combining spectral, TIN-based structural and meteorological features. *Remote Sensing*, *11*(7). https://doi.org/10.3390/RS11070890

Jiang, Y., Li, C., Robertson, J. S., Sun, S., Xu, R., & Paterson, A. H. (2018). GPhenoVision: A ground mobile system with multi-modal imaging for field-based high throughput

phenotyping of cotton. *Scientific Reports*, *8*(1), 1–15. https://doi.org/10.1038/s41598-018-19142-2

Johnson, R. M., & Richard, E. P. (2010). Variable-rate lime application in Louisiana sugarcane production systems. *Precision Agriculture*, *11*(5), 464–474. https://doi.org/10.1007/s11119-009-9140-2

Keating, B. A., Robertson, M. J., Muchow, R. C., & Huth, N. I. (1999). Modelling sugarcane production systems I. Development and performance of the sugarcane module. *Field Crops Research*, *61*(3), 253–271. https://doi.org/10.1016/S0378-4290(98)00167-1

Khan, M. T., & Khan, A. (2019). *Sugarcane Biofuels* (M. T. Khan & I. A. Khan, Eds.). https://doi.org/10.1007/978-3-030-18597-8

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., & Shearer, S. (2018). Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, *153*(July), 213–225. https://doi.org/10.1016/j.compag.2018.07.016

Kim, M., Ko, J., Jeong, S., Yeom, J., Kim, H., Kim, M., … Kim, H. (2017). Monitoring canopy growth and grain yield of paddy rice in South Korea by using the GRAMI model and high spatial resolution imagery. *GIScience & Remote Sensing*, *54*(4), 534–551. https://doi.org/10.1080/15481603.2017.1291783

Kim, N., Na, S. Il, Park, C. W., Huh, M., Oh, J., Ha, K. J., … Lee, Y. W. (2020). An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data. *Applied Sciences (Switzerland)*, *10*(11). https://doi.org/10.3390/app10113785

Kotchenova, S. Y., Vermote, E. F., Matarrese, R., & Klemm, F. J. (2006). Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. *Applied Optics*, *45*(26), 6762–6774. https://doi.org/10.1364/AO.45.006762

Krishnan, P., Ramakrishnan, B., Reddy, K. R., & Reddy, V. R. (2011). High-Temperature Effects on Rice Growth, Yield, and Grain Quality. In *Advances in Agronomy* (1st ed., Vol. 111). https://doi.org/10.1016/B978-0-12-387689-8.00004-7

Lee, C. (1960). *A Culture History of Rice With Special Reference to Louisiana* .

Leroux, C., Jones, H., Clenet, A., Dreux, B., Becu, M., & Tisseyre, B. (2018). A general method to filter out defective spatial observations from yield mapping datasets. *Precision Agriculture*, *19*(5), 789–808. https://doi.org/10.1007/s11119-017-9555-0

Linsley, C. M., & Bauer, F. C. (1929). Test your soil for acidity. *University of Illinois, Agricultural Experiment Station ; No. 346*, 2–4.

Lofton, J., & Tubaña, B. (2016). *Effect of Nitrogen Rates and Application Time on Sugarcane Yield and Quality*. *4167*(October). https://doi.org/10.1080/01904167.2013.828752

Lofton, J., Tubana, B. S., Kanke, Y., Teboh, J., Viator, H., & Dalen, M. (2012). Estimating sugarcane yield potential using an in-season determination of normalized difference vegetative index. *Sensors*, *12*(6), 7529–7547. https://doi.org/10.3390/s120607529

Louisiana Rice Research Board, & LSU AgCenter. (2018). *Louisiana Rice Research Board Annual Report*.

Lowenberg-DeBoer, J., & Erickson, B. (2019). Setting the Record Straight on Precision Agriculture Adoption. *Agronomy Journal*, *111*(4), 1552–1569. https://doi.org/10.2134/agronj2018.12.0779

LSU AgCenter. (2018). *Louisiana Summary: Agriculture & Natural Resources*. Retrieved from https://www.lsuagcenter.com/profiles/aiverson/articles/page1570114750015

Lyle, G., Bryan, B. A., & Ostendorf, B. (2014). Post-processing methods to eliminate erroneous grain yield measurements: Review and directions for future development. *Precision Agriculture*, *15*(4), 377–402. https://doi.org/10.1007/s11119-013-9336-3

Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A. M., Erkbol, H., & Fritschi, F. B. (2020). Crop monitoring using satellite/UAV data fusion and machine learning. *Remote Sensing*, *12*(9). https://doi.org/10.3390/RS12091357

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, *237*(December 2019), 111599. https://doi.org/10.1016/j.rse.2019.111599

Maldaner, L F, Corrêdo, L. P., Tavares, T. R., Mendez, L. G., Duarte, C., & Molin, J. P. (2018). Identifying and filtering out outliers in spatial datasets. *Proceedings of the 14th International Conference on Precision Agriculture*, unpaginated, online. Monticello, IL.

Maldaner, Leonardo Felipe, Molin, J. P., & Canata, T. F. (2016). Processing yield data from two or more combines. *Proceedings of the 13th International Conference on Precision Agriculture*, (September).

Marin, F. R., & Jones, J. W. (2014). Process-based simple model for simulating sugarcane growth and production. *Scientia Agricola*, *71*(1), 1–16. https://doi.org/10.1590/S0103-90162014000100001

Martin, K., Raun, W., & Solie, J. (2012). By-plant prediction of corn grain yield using optical sensor readings and measured plant height. *Journal of Plant Nutrition*, *35*(9), 1429–1439. https://doi.org/10.1080/01904167.2012.684133

Mcmaster, G. S., Wilhelm, W., & Wilhelm, W. W. (1997). Growing degree-days: one equation, two interpretations. *Agricultural and Forest Meteorology*, *87*, 291–300. Retrieved from https://digitalcommons.unl.edu/usdaarsfacpubhttps://digitalcommons.unl.edu/usdaarsfacpub/83

Menegatti, L. A. A., & Molin, J. P. (2004). Removal of errors in yield maps through raw data filtering. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *8*(1), 126–134. https://doi.org/10.1590/s1415-43662004000100019

Molijn, R. A., Iannini, L., Rocha, J. V., & Hanssen, R. F. (2019). Sugarcane productivity mapping through C-band and L-band SAR and optical satellite imagery. *Remote Sensing*, *11*(9), 1–27. https://doi.org/10.3390/rs11091109

Molin, J. P., & Tavares, T. R. (2019). Sensor Systems for Mapping Soil Fertility Attributes: Challenges, Advances, and Perspectives in Brazilian Tropical Soils. *Engenharia Agrícola*, *39*(spe), 126–147. https://doi.org/10.1590/1809-4430-eng.agric.v39nep126-147/2019

Momin, M. A., Grift, T. E., Valente, D. S., & Hansen, A. C. (2018). Sugarcane yield mapping based on vehicle tracking. *Precision Agriculture*, (0123456789). https://doi.org/10.1007/s11119-018-9621-2

Morel, J., Bégué, A., Todoroff, P., Martiné, J. F., Lebourgeois, V., & Petit, M. (2014). Coupling a sugarcane crop model with the remotely sensed time series of fIPAR to optimise the yield estimation. *European Journal of Agronomy*. https://doi.org/10.1016/j.eja.2014.08.004

Nevavuori, P., Narra, N., & Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Computers and Electronics in Agriculture*, *163*(June), 104859. https://doi.org/10.1016/j.compag.2019.104859

Pallottino, F., Antonucci, F., Costa, C., Bisaglia, C., Figorilli, S., & Menesatti, P. (2019). Optoelectronic proximal sensing vehicle-mounted technologies in precision agriculture : A review Partial Least Squares Regression. *Computers and Electronics in Agriculture*, *162*(May), 859–873. https://doi.org/10.1016/j.compag.2019.05.034

Passalaqua, B. P., & Molin, J. P. (2020). Path Errors in Sugarcane Transshipment Trailers. *Engenharia Agrícola*, *40*(2), 223–231. https://doi.org/10.1590/1809-4430-eng.agric.v40n2p223-231/2020

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439. https://doi.org/10.32614/RJ-2018-009

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. Retrieved from http://arxiv.org/abs/1201.0490

Peerlinck, A., Sheppard, J., & Maxwell, B. (2018). *Using Deep Learning in Yield and Protein Prediction of Winter Wheat Based on Fertilization Prescriptions in Precision Agriculture A paper from the Proceedings of the 14 th International Conference on Precision Agriculture*. 1–13.

Peralta, N. R., Assefa, Y., Du, J., Barden, C. J., & Ciampitti, I. A. (2016). Mid-season high-resolution satellite imagery for forecasting site-specific corn yield. *Remote Sensing*, *8*(10). https://doi.org/10.3390/rs8100848

Peteinatos, G. G., Weis, M., Andújar, D., Rueda Ayala, V., & Gerhards, R. (2014). Potential use of ground-based sensor technologies for weed detection. *Pest Management Science*, *70*(2), 190–199. https://doi.org/10.1002/ps.3677

Planet Labs Inc. (2020). *Planet Imagery Product Specifications*. Retrieved from https://assets.planet.com/docs/combined-imagery-product-spec-final-august-2019.pdf

Planet Team. (2018). *Planet Application Program Interface: In Space for Life on Earth*. Retrieved from https://api.planet.com

Portz, G., Amaral, L. R., Molin, J. P., & Jasper, J. (2012). Optimum Sugarcane Growth Stage for Canopy Reflectance Sensor to Predict Biomass and Nitrogen Uptake. *16th International Conference on Precision Agriculture*.

Prichard, W. (1939). The Effects of the Civil War on the Louisiana Sugar Industry. *The Journal of Southern History*, *5*(3), 315–332.

Purcell, L. C. (2003). COMPARISON OF THERMAL UNITS DERIVED FROM DAILY AND HOURLY TEMPERATURES. *Crop Science*, *43*(5), 1874–1879. https://doi.org/10.2135/cropsci2003.1874

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Rahman, M. M., & Robson, A. (2020). Integrating landsat-8 and sentinel-2 time series data for yield prediction of sugarcane crops at the block level. *Remote Sensing*, *12*(8), 1–15. https://doi.org/10.3390/RS12081313

Ramos, A. P. M., Osco, L. P., Furuya, D. E. G., Gonçalves, W. N., Santana, D. C., Teodoro, L. P. R., … Pistori, H. (2020). A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, *178*(July), 105791. https://doi.org/10.1016/j.compag.2020.105791

Raper, T. B., Varco, J. J., & Hubbard, K. J. (2013). Canopy-based normalized difference vegetation index sensors for monitoring cotton nitrogen status. *Agronomy Journal*, *105*(5), 1345–1354. https://doi.org/10.2134/agronj2013.0080

Rischbeck, P., Elsayed, S., Mistele, B., Barmeier, G., Heil, K., & Schmidhalter, U. (2016). Data fusion of spectral, thermal and canopy height parameters for improved yield prediction of drought stressed spring barley. *European Journal of Agronomy*, *78*, 44–59. https://doi.org/10.1016/j.eja.2016.04.013

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems

in the great plains with ERTS. *Third Earth Resources Technology Satellite (ERTS) Symposium*, *1*, 309–317. https://doi.org/citeulike-article-id:12009708

Ruß, G., & Brenning, A. (2010). Data Mining in Precision Agriculture: Management of Spatial Information. In H. E., K. R., & H. F. (Eds.), *Computational Intelligence for Knowledge-Based Systems Design. IPMU 2010. Lecture Notes in Computer Science, vol 6178* (pp. 350–359). https://doi.org/10.1007/978-3-642-14049-5_36

Saichuk, J. K., Hollier, C. A., White, L. M., Webster, E. P., Reagan, T. E., Schultz, B., … Linscombe, S. D. (2014). *Louisiana Rice Production Handbook*.

Sanches, G. M., Duft, D. G., Kölln, O. T., Luciano, A. C. dos S., De Castro, S. G. Q., Okuno, F. M., & Franco, H. C. J. (2018). The potential for RGB images obtained using unmanned aerial vehicle to assess and predict yield in sugarcane fields. *International Journal of Remote Sensing*, *39*(15–16), 5402–5414. https://doi.org/10.1080/01431161.2018.1448484

Schueller, J. K., Whitney, J. D., Wheaton, T. A., Miller, W. M., & Turner, A. E. (1999). Low-cost automatic yield mapping in hand-harvested citrus. *Computers and Electronics in Agriculture*, *23*(2), 145–153. https://doi.org/10.1016/S0168-1699(99)00028-9

Schut, A. G. T., Traore, P. C. S., Blaes, X., & de By, R. A. (2018). Assessing yield and fertilizer response in heterogeneous smallholder fields with UAVs and satellites. *Field Crops Research*, *221*(June 2017), 98–107. https://doi.org/10.1016/j.fcr.2018.02.018

Scott, A. J., & Knott, M. (1974). A Cluster Analysis Method for Grouping Means in the Analysis of Variance. *Biometrics*, *30*(3), 507. https://doi.org/10.2307/2529204

Serrano, J., Shahidian, S., & da Silva, J. M. (2014). Spatial and temporal patterns of apparent electrical conductivity: DUALEM vs. Veris sensors for monitoring soil properties. *Sensors (Switzerland)*, *14*, 10024–10041. https://doi.org/10.3390/s140610024

Shannon, D. K., Clay, D. E., & Kitchen, N. R. (2018). *Precision Agriculture Basics*. https://doi.org/10.2134/precisionagbasics

Shiratsuchi, L., Ferguson, R., Shanahan, J., Adamchuk, V., Rundquist, D., Marx, D., & Slater, G. (2011). Water and nitrogen effects on active canopy sensor vegetation indices. *Agronomy Journal*, *103*(6), 1815–1826. https://doi.org/10.2134/agronj2011.0199

Silva, C. B., de Moraes, M. A. F. D., & Molin, J. P. (2011). Adoption and use of precision agriculture technologies in the sugarcane industry of São Paulo state, Brazil. *Precision Agriculture*, *12*(1), 67–81. https://doi.org/10.1007/s11119-009-9155-8

Sofonia, J., Shendryk, Y., Phinn, S., Roelfsema, C., Kendoul, F., & Skocaj, D. (2019). Monitoring sugarcane growth response to varying nitrogen application rates: A comparison of UAV SLAM LiDAR and photogrammetry. *International Journal of Applied Earth Observation and Geoinformation*, *82*(June), 101878. https://doi.org/10.1016/j.jag.2019.05.011

Solie, J. B., Monroe, A. D., Raun, W. R., & Stone, M. L. (2012). Generalized Algorithm for

Variable-Rate Nitrogen Application in Cereal Grains. *Agronomy Journal*, *104*(2), 378–387. https://doi.org/10.2134/agronj2011.0249

Son, N. T., Chen, C. F., Chen, C. R., Chang, L. Y., Duc, H. N., & Nguyen, L. D. (2013). Prediction of rice crop yield using MODIS EVI-LAI data in the Mekong Delta, Vietnam. *International Journal of Remote Sensing*, *34*(20), 7275–7292. https://doi.org/10.1080/01431161.2013.818258

Spekken, M., Anselmi, A. A., & Molin, J. P. (2013). A simple method for filtering spatial data. In J. V. Stafford (Ed.), *Precision agriculture'13: Proceedings of the 9th European conference on precision agriculture* (pp. 259–266). Wageningen: The Netherlands: Wageningen Academic Publishers.

SSS, NCRS, & USDA. (2020). Soil Survey Geographic (SSURGO) Database. Retrieved from https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx

Sudduth, K. A., & Drummond, S. T. (2007). Yield Editor: Software for Removing Errors from Crop Yield Maps. *Agronomy Journal*, *99*(6), 1471–1482. https://doi.org/10.2134/agronj2006.0326

Sudduth, K. A., Drummond, S. T., & Myers, D. B. (2012). Yield editor 2.0: Software for automated removal of yield map errors. *American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012*, *4*(12), 3378–3391. https://doi.org/10.13031/2013.41893

Sui, R., Thomasson, J. A., Hanks, J., & Wooten, J. (2008). Ground-based sensing system for weed mapping in cotton. *Computers and Electronics in Agriculture*, *60*(1), 31–38. https://doi.org/10.1016/j.compag.2007.06.002

Sun, J., Di, L., Sun, Z., Shen, Y., & Lai, Z. (2019). County-level soybean yield prediction using deep CNN-LSTM model. *Sensors (Switzerland)*, *19*(20), 1–21. https://doi.org/10.3390/s19204363

Tangwongkit, R. (2006). Development of a real-time, variable rate herbicide applicator using machine vision for between-row weeding of sugarcane fields. *...International: The CIGR ...*, *VIII*, 6009. Retrieved from http://cigrjournal.org/index.php/Ejounral/article/view/659/653

Tavares, T. R., Molin, J. P., Nunes, L. C., Alves, E. E. N., Melquiades, F. L., de Carvalho, H. W. P., & Mouazen, A. M. (2020). Effect of X-Ray Tube Configuration on Measurement of Key Soil Fertility Attributes with XRF. *Remote Sensing*, *12*(6), 963. https://doi.org/10.3390/rs12060963

Tennekes, M. (2018). {tmap}: Thematic Maps in {R}. *Journal of Statistical Software*, *84*(6), 1–39. https://doi.org/10.18637/jss.v084.i06

Teruel, D. A., Barbieri, V., & Ferraro Jr., L. A. (1997). Sugarcane leaf area index modeling under different soil water conditions. *Scientia Agricola*, *54*(spe), 39–44.

https://doi.org/10.1590/S0103-90161997000300008

Trudgill, D. L., Honek, A., Li, D., & Straalen, N. M. (2005). Thermal time - concepts and utility. *Annals of Applied Biology*, *146*(1), 1–14. https://doi.org/10.1111/j.1744-7348.2005.04088.x

Tukey, J. W. (1977). *Exploratory Data Analysis*. Mass: Addison-Wesley Pub. Co.

USDA. (2019). Tailored Reports: Crop Production Practices. Retrieved October 14, 2020, from https://data.ers.usda.gov/reports.aspx?ID=17883

Vega, A., Córdoba, M., Castro-Franco, M., & Balzarini, M. (2019). Protocol for automating error removal from yield maps. *Precision Agriculture*, *20*(5), 1030–1044. https://doi.org/10.1007/s11119-018-09632-8

Wan, L., Cen, H., Zhu, J., Zhang, J., Zhu, Y., Sun, D., … He, Y. (2020). Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer – a case study of small farmlands in the South of China. *Agricultural and Forest Meteorology*, *291*(March), 108096. https://doi.org/10.1016/j.agrformet.2020.108096

Wei, M. C. F., Maldaner, L. F., Ottoni, P. M. N., & Molin, J. P. (2020). Carrot Yield Mapping: A Precision Agriculture Approach Based on Machine Learning. *Artificial Inteligence*, *1*(2), 229–241. https://doi.org/10.3390/ai1020015

Yang, C., Everitt, J. H., Du, Q., Luo, B., & Chanussot, J. (2013). Using high-resolution airborne and satellite imagery to assess crop growth and yield variability for precision agriculture. *Proceedings of the IEEE*, *101*(3), 582–592. https://doi.org/10.1109/JPROC.2012.2196249

Yang, M., Yang, J., Su, L., Sun, K., Li, D., Liu, Y., … Guo, T. (2019). Metabolic profile analysis and identification of key metabolites during rice seed germination under low-temperature stress. *Plant Science*, *289*(July), 110282. https://doi.org/10.1016/j.plantsci.2019.110282

Yao, Y., Miao, Y., Huang, S., Gao, L., Chen, X., Zhang, F., & Yu, K. (2012). *Active canopy sensor-based precision N management strategy for rice*. *2006*, 925–933. https://doi.org/10.1007/s13593-012-0094-9

Zhao, G., Miao, Y., Wang, H., Su, M., Fan, M., Zhang, F., … Ma, D. (2013). A preliminary precision rice management system for increasing both grain yield and nitrogen use efficiency. *Field Crops Research*, *154*, 23–30. https://doi.org/10.1016/j.fcr.2013.07.019

Zhao, Q., Lenz-Wiedemann, V. I. S., Yuan, F., Jiang, R., Miao, Y., Zhang, F., & Bareth, G. (2015). Investigating within-field variability of rice from high resolution satellite imagery in Qixing farm county, northeast China. *ISPRS International Journal of Geo-Information*, *4*(1), 236–261. https://doi.org/10.3390/ijgi4010236

Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B., & Hammer, G. L. (2020). Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sensing*, *12*(6), 1024. https://doi.org/10.3390/rs12061024

Vita

Felippe H. S. Karp was born in São João da Boa Vista, Brazil in March of 1996. He attended to the University of São Paulo campus College of Agriculture “Luiz de Queiroz” and received his Bachelor of Science in Agronomic Engineer in December of 2018. After graduating he came to Louisiana State University to pursue his Master’s degree under the School of Plant Environmental and Soil Science program. He developed his Master’s thesis under the supervision of Dr. Luciano Shiratsuchi focusing on the evaluation of yield map prediction models based on high-resolution satellite imagery and machine learning techniques with the purpose of supporting the development of precision agriculture in different crops, specially sugarcane and rice. He plans to receive his Master’s this December 2020 and move forward to continue his studies pursuing a PhD.